DOCUMENT RESUME

ED 061 065 SE 013 397

TITLE	Evaluation in Chemistry. Report of International Workshop, Ceylon, August 1968.
INSTITUTION	International Union of Pure and Applied Physics.; United Nations Educational, Scientific, and Cultural Organization, Paris (France).
PUB DATE	Aug 68
NOTE	165p.
EDRS PRICE	MF-$0.65 HC-$6.58
DESCRIPTORS	*Chemistry; College Science; Conference Reports; *Curriculum Development; Evaluation; *International Education; National Programs; Science Education; Secondary School Science; *Testing
IDENTIFIERS	UNESCO

ABSTRACT

Included are the proceedings and conclusions of an international workshop held in Ceylon in August, 1968, which was convened by the International Union of Pure and Applied Chemistry (IUPAC). The report is concerned not only with the design, construction and administration of chemistry examinations and evaluation of chemistry curricula but with the whole interrelated problem of curriculum change within an educational system. The first part of the report considers: (1) Aims and Objectives of Education through Chemistry, (2) The Functions of Chemistry Examinations in an Educational System, (3) Administrative and Teaching Action Necessary to Achieve Objectives, (4) Planning an Examination Scheme, (5) Administration and Execution of Chemistry Examinations, and (6) Strategy and Tactics for Curriculum Reform. The rest of the report consists of supporting evidence and background papers from actual developments in Ceylon, India, United Kingdom, and the Soviet Union. (Author/TS)

ED 061065

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

EVALUATION IN CHEMISTRY

REPORT OF INTERNATIONAL WORKSHOP

CEYLON AUGUST 1968

SE 013 397

IUPAC UNESCO

EVALUATION IN CHEMISTRY

Report of a workshop sponsored by IUPAC and UNESCO
and held in the University of Ceylon Peradeniya
13th-18th August 1968

FOREWORD

When the International Union of Pure and Applied Chemistry (IUPAC) established in 1965 a Committee on the Teaching of Chemistry, one of the first tasks of this Committee was to review major areas of chemical education and decide on priorities for action. It was quickly agreed that examinations in chemistry play a major part in determining how the subject is taught and that if new approaches to teaching chemistry were desired, then one way of bringing this about would be to change the type of questions asked in public and internal examinations.

Accordingly the IUPAC Committee on the Teaching of Chemistry invited Mr. J.C. Mathews (U.K.) to prepare a report on "The Effect of Examinations in Determining Chemistry Curricula up to University Level" which was published in Pure and Applied Chemistry by Butterworths in 1967. This Report was mainly a comparative survey of selected countries in Europe together with U.S.A. and Japan.

Subsequently a report was submitted to the IUPAC Committee on similar problems viewed from the practice in France and countries in Southern Europe, whose educational systems and examination procedures are somewhat different from Western Europe and the U.S.A. This report has not yet been published.

The Report that follows in this booklet describes the proceedings and conclusions of an international workshop held in Ceylon in August, 1968, which was convened by IUPAC as a follow-up to the original Mathews Report. The Workshop was under the Chairmanship of Professor H.F. Halliwell (U.K.), who has had wide experience in chemical education throughout the world, and the Secretary to the Workshop was Mr. D.G. Chisman, Secretary of IUPAC Committee on the Teaching of Chemistry. Mr. J.C. Mathews himself was among the ten participants from outside Ceylon.

The Workshop was held in Ceylon at the invitation of the Ceylonese Ministry of Education and was timed to follow a seminar on evaluation and curriculum development sponsored by the Ministry of Education in Ceylon, the British Ministry of Overseas Development and British Council which was under the direction of Mr. J.C. Mathews. The IUPAC Committee on the Teaching of Chemistry were pleased to have this opportunity of arranging a Workshop in Ceylon on this topic since the progress made by the Ceylon Ministry of Education in revising chemistry curricula and changing the whole approach to examining in chemistry is very considerable indeed. They were also pleased to be able to collaborate with UNESCO in the joint sponsorship of the Workshop on this occasion since UNESCO's Chemistry Pilot Project in Asia has been involved with similar problems of evaluation. In particular the UNESCO Study Group in Ceylon, associated with the Pilot Project was involved in the Workshop.

ACKNOWLEDGMENTS

The IUPAC Committee on the Teaching of Chemistry wish to acknowledge with grateful thanks the help and financial assistance provided by:

The Ministry of Education in Ceylon

University of Ceylon, Peradeniya

UNESCO

National Science Foundation, U.S.A.

American Chemical Society

Centre for Curriculum Renewal and Educational Development Overseas (CREDO), U.K.

British Council

Ministry of Overseas Development, U.K.

Special thanks are due to Mr J.P. Alles, Deputy Director General (Secondary Education), Ministry of Education, Ceylon, and to Mr. J. Ratnaike, Chief Education Officer (Curriculum), Science/Mathematics Project, Ministry of Education, Ceylon, and to their many colleagues whose hard work before and during the meeting in preparing and editing Working Papers, Resource Material and background material contributed greatly to the success of the Workshop.

Report of a workshop sponsored by IUPAC and UNESCO
and held in the University of Ceylon Peradeniya
13th-18th August 1968

Page

Foreword

Acknowledgments

Introduction 1

Part I - The Main Report

1. Aims and Objectives of Education through Chemistry. 2

2. The Functions of Chemistry Examinations in an Educational System. 5

3. Administrative and Teaching Action Necessary to Achieve Objectives. 8

4. Planning an Examination Scheme. 10

5. Administration and Execution of Chemistry Examinations. 16

6. Strategy and Tactics for Curriculum Reform. 37

Appendix A Glossary 39

Appendix B Bibliography 42

Appendix C List of participants 44

Part II - Supporting Evidence

1. The Educational System of Ceylon, from the point of view of science education (including Table of Specifications), by H.F. Halliwell. 46

2. Design for Setting Sample Question Paper in Chemistry, NCERT India (including blueprint of specifications). 64

3. Notes on the Assessment of Practical Work: 'A' Level Chemistry, Nuffield Scheme, U.K., by J.C. Mathews. 69

Part III - Background Papers

		Page
1.	Evaluation in Chemistry within a Second Level General Education Programme, by J.P. Alles.	80
2.	Evaluation - A Critical Stimulus to Science Curriculum Development in Ceylon, by J. Ratnaike.	91
3.	Improvement of Evaluation Practices, by P. Samaranayaka.	100
4.	"Notes on Structural and Functional Aspects of an Educational System" (extract from pp. 23-28 of Ministry of Education Ceylon publication).	109
5.	Oral Evaluation, with Particular Reference to Soviet Schools, by S.A. Balezin.	113
6.	Report on Evaluation Workshop in Chemistry: Peradeniya, Ceylon, 2-12 August 1968, by J.C. Mathews and J. Ratnaike.	118

INTRODUCTION

In order to communicate clearly and effectively in any systematic activity it becomes necessary to develop concepts and at the same time to evolve symbolism and technical terms which carry specific and unambiguous meaning.

Chemistry as a subject provides a good example of the need to develop such concepts with associated symbolisms and technical terminology. The emergence of chemical, symbolic nomenclature in the early 19th century contributed significantly to the growth of chemistry. In recent years, the study of organic reaction mechanism and the concepts evolved for its interpretation and communication provides a further example.

In the field of curriculum development and assessment the use of specialised terminology and symbolism, such as content objectives, behavioural objectives, tables of specifications, reliability coefficient, difficulty indices and discrimination coefficients, illustrate an analogous phase of professional activity.

In this report terms that are commonly used in curriculum development have been included, and there is a glossary of such terms for those new to this field. It is suggested, however, that those chemists and administrators who are now engaged in this important area of evaluation and assessment of chemists curricula should become sensitive to the concepts involved and familiar in the use of related symbolism and terminology. This will lead not only to an enhanced lucidity in communication among professional workers but will also provide for an accelerated development of chemical education itself.

Before the specific problems of evaluation of chemistry curricula can be considered it is necessary to place chemistry in perspective within an educational system, to reflect on the aims and objectives of education as a whole and the part that chemistry as a subject can play in achieving such aims. Thus the report that follows is concerned not only with the design, construction and administration of chemistry examinations and evaluation of chemistry curricula but with the whole interrelated problem of curriculum change within an educational system.

I THE MAIN REPORT

1. Aims and Objectives of Education through Chemistry

The aims and objectives of an educational system may be regarded as part of a cyclic process of curriculum development as indicated in Figure 1.

Objectives may be seen and understood differently by pupils, teachers, curriculum designers and policy makers. Since this report is concerned essentially with evaluation, objectives will be considered mainly from the point of view of those involved in curriculum assessment. In this context it is possible to outline broad classifications of objectives and this may be preferable to listing detailed objectives. Such amplification in any case must relate to the pupils for whom the curriculum is intended and adjusted to the conditions prevailing in a particular society.

Objectives may be broadly classified in two divisions - content objectives and behavioural objectives.

The content objectives will vary with time and place as the subject develops and grows. The design of chemistry courses will have to reflect this change and the content should include carefully selected themes from chemistry to enable pupils to acquire an ordered view of nature and materials. It should also give pupils an appreciation of chemistry in everyday life, indicate the results of the application of chemistry to technology, agriculture and medicine, and show the relationship of chemistry to other sciences and culture. These themes must be treated in such a way as to give some awareness of the role of chemistry in terms of individual and social needs and of the scientific approach to problems.

The behavioural objectives relate to changes which chemistry teaching seeks to achieve in pupils' skills, attitudes and knowledge. These changes will be brought about primarily through pupil participation in experimental work leading to the development of three aspects of learning. First, the ability to use simple apparatus to observe changes in materials in chemical reactions and to record such changes in a meaningful way. Secondly, the ability to recall, classify and interpret information and experiences, to devise appropriate schemes for solving a practical problem, to apply previous understanding to new situations and to report and comment on matters of

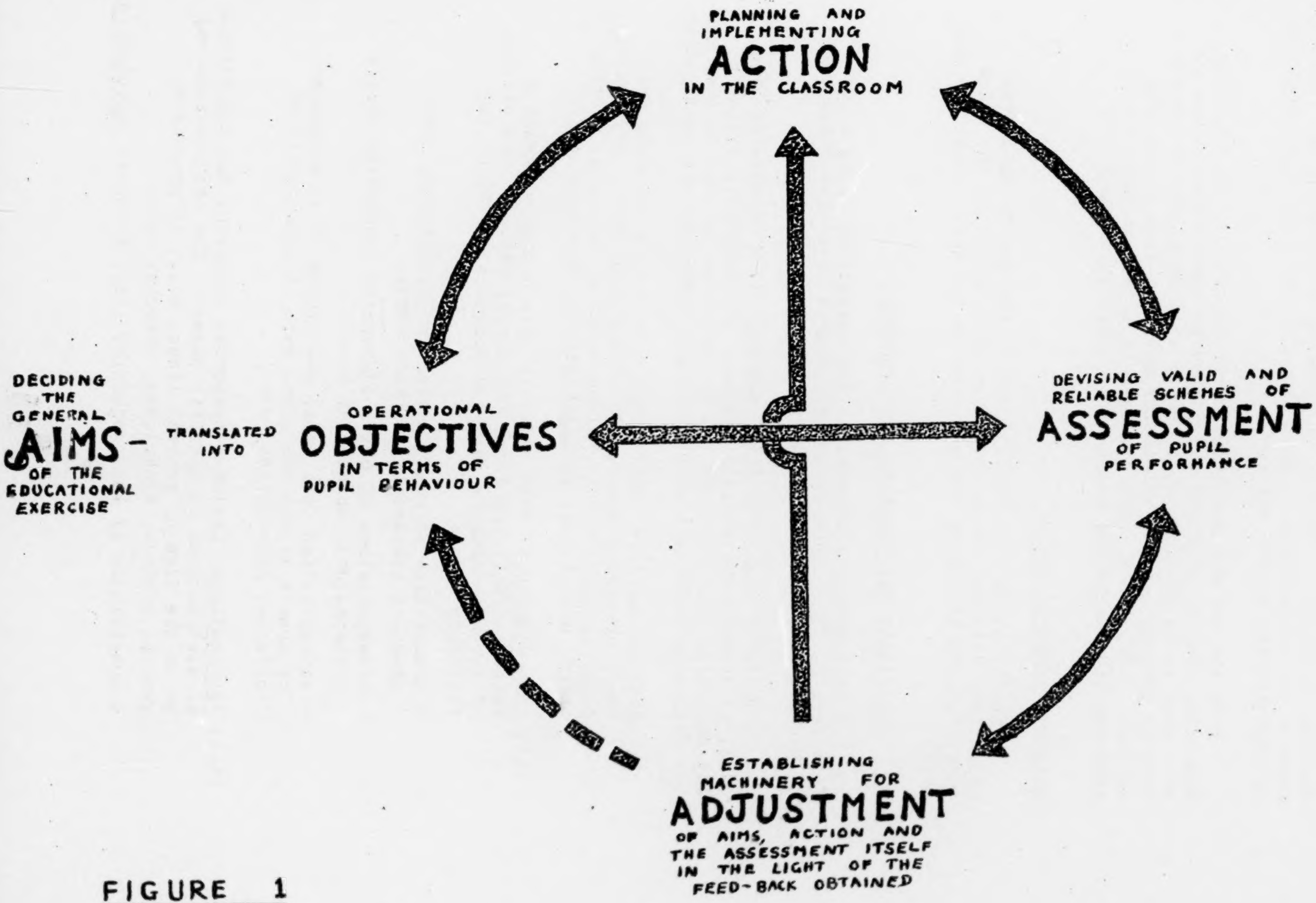

PLANNING AND IMPLEMENTING
ACTION
IN THE CLASSROOM
DECIDING THE GENERAL
AIMS -
OF THE EDUCATIONAL EXERCISE
TRANSLATED INTO
OPERATIONAL
OBJECTIVES
IN TERMS OF PUPIL BEHAVIOUR
DEVISING VALID AND RELIABLE SCHEMES OF
ASSESSMENT
OF PUPIL PERFORMANCE
ESTABLISHING MACHINERY FOR
ADJUSTMENT
OF AIMS, ACTION AND THE ASSESSMENT ITSELF IN THE LIGHT OF THE FEED-BACK OBTAINED

FIGURE 1

chemical interest. Thirdly, to develop an interest and involvement in the subject and to gain satisfaction from it, leading to an appreciation of its value.

Both the content and behavioural objectives of teaching chemistry need to be precisely stated in operational terms and will need to be modified accordingly when chemistry is considered as part of general science as distinct from a separate subject. They must, however, seek to establish securely in students the desire and capacity to learn throughout life.

Interpretation in Ceylon

The following examples illustrate the way in which the objectives outlined above have been interpreted in the educational system in Ceylon and conveyed to teachers, examiners and pupils.

a) Objectives for Teachers and Examiners

In this analysis objectives are classified using a modified form of "The Taxonomy of Educational Objectives" designed by Bloom et al.(see bibliography).

A Table of Specifications (see Part II - Supporting Evidence, 1) listing the various items of the chemistry syllabus and analysing the questions set on each item is used to show how the following different objectives are tested in the examination.

(i) Recall: involves test items requiring a response utilising primarily the recall of specific parts, terminology, conversations, trends, classification criteria, methodology, principles, generalisations, theories, etc.

(ii) Comprehension: test items in this category demand a response requiring more than simple recall. An element of understanding is required and this may be of the following types:

- translation of mathematical verbal material into symbolic statements and vice versa,
- interpretation of data, statements (including simple explanations, summaries, etc.),
- extrapolation from data, statements (i.e. extension of trends beyond the given data, determination of effects, corollaries, etc.).

(iii) Application: includes responses involving the application of abstractions in specific cases. The abstractions may be in the form of general ideas, rules of procedure, general methods, techniques, theories, etc.

A subdivision of this objective (iii) involves calculation.

(iv) Analysis, synthesis and evaluation: includes items testing recognition of explicit and implicit relationships, structure; recognition of unstated assumptions; discriminations of facts from hypotheses; organisation of ideas, statements, data; design of simple experiments; recognition of logical accuracy, consistency; comparison of theories and generalisations.

b) Instructions to Schools: O-level Chemistry Examinations

In the introductory notes to the new O-level chemistry syllabus (see bibliography) candidates are informed that there will be two papers - one of three quarters of an hour duration consisting of short answer questions carrying 40% of the marks, the other of two hour duration. They are also advised that in their answers they will be expected to give evidence of:

(i) An understanding of the facts, principles and theories of elementary chemistry.

(ii) An appreciation and an understanding of the relevance of elementary chemical facts, principles and theories to everyday life.

(iii) A simple appreciation and an understanding of the methods of experimental science as applied in elementary chemistry.

(iv) A familiarity with simple apparatus and experimental work. (The possibility of using simple materials, other than standard laboratory apparatus, to perform experiments, should be recognized. Teachers are advised to use such materials wherever they are relevant.)

2. The Function of Examinations in Chemistry in an Educational System

As outlined in the previous chapter the process of curriculum development may be regarded in a cyclic manner (Fig. 1). Having formulated aims and objectives, appropriate action is required to achieve these goals. In turn assessment is used to determine to what extent and how successfully the aims and objectives have been achieved and to adjust both the objectives and action if necessary.

Thus examinations as one means of assessment play a powerful and somewhat complex role in the development of an educational system. In various systems examinations in chemistry are currently being used for some or all of the following purposes. It is unlikely, however, that any one examination would be valid for all the purposes listed.

The first three purposes listed are directed at the pupils; the remainder are aimed at the educational system as a whole. In a well established system the main purposes of examinations may be simply the first three, but in a developing system the other purposes can be of great importance in providing data on which judgments and adjustments can be made. In particular (vi) and (viii) can have a major formative influence on teaching and learning.

(i) To test achievement of pupils.

(ii) To indicate potential of pupils.

(iii) As a selection instrument for higher education or employment.

(iv) To assess effectiveness of teaching.

(v) To diagnose pupils' difficulties.

(vi) To stimulate and modify teaching and learning methods and attitudes.

(vii) To modify future examinations.

(viii) To spell out the objectives and content of the curriculum in operational and specific terms.

(ix) To act as a sensitive and rapid steering system to adjust decision making at the aims and objectives level.

(x) To reflect quantitatively the characteristics of an educational environment.

(xi) To locate constraints in an educational system.

Examinations in Ceylon

Chemistry examinations in the Ceylon educational system are used for all the purposes listed above in the following way.

The Ceylon G.C.E. 'O' and 'A' level examinations are used as achievement tests and also for selection purposes ((i) and (iii)). In the absence of suitable I.Q. or aptitude tests the G.C.E. examination in science, mathematics and first language is also used for predicting the potential of pupils ((ii)) using correlations from the G.C.E. National Stratified Random Sample Data (see Part III - Background Papers 2 and 3).

The effectiveness of teaching ((iv)) particular areas of the syllabus and teacher performance in general are measured by Difficulty Indices and Means from the G.C.E. 'O' National Stratified Random Data. Likewise specific points in the teaching scheme are used for diagnostic purposes ((v)) - to indicate pupils' difficulties and to suggest reinforcement and new teaching sequences. In-service training courses are aimed at specific areas of the syllabus as a corrective to weaknesses brought to light in the data.

During the period 1961-65 changes in the examinations in chemistry were designed to stimulate changes in the teaching ((vi)). A greater coverage of the old syllabus was made possible by changing the examination structures and also objectives such as application, analysis and synthesis (see Part III - Background Paper 2) were introduced. This period of change made it possible to introduce a major curriculum and examination reform in 1965. Further changes in the design and moderation of chemistry examinations ((vii)) are brought about by reference to the National Stratified Random Sample. The teachers and pupils may refer to a prototype paper and marking schemes attached to the new syllabus and teaching schemes in order to become completely familiar with the objectives and content of the curriculum in specific terms ((viii)).

The way in which an examination can be used as a sensitive and rapid steering instrument ((ix)) to assist policy making may be exemplified by the problem facing Ceylon in the mid-1960s in the wish to extend science teaching facilities in rural areas but with a severe limit on the number of university science graduate teachers available. Ordinary trained (non-graduate) teachers were available and some were teaching at G.C.E. 'O' level but doubt on their suitability for such work was widely expressed. Data from the National Stratified Random Sample for December 1965 and 1966 were studied and it was shown that these doubts were not reflected in the achievement of pupils. Thus, in this specific context, the continued deployment of non-graduate teachers at this level was regarded as permissible as an emergency measure.

Data from the G.C.E. 'O' examinations 1952-1958 were used to demonstrate the quantitative relationship of performance with educational environment and the more recent data from the National Random Stratified Sample are being used to measure constraints on the educational system such as urban/rural differences, socio-economic and educational status of parents, as well as teacher qualification and experience ((x) and (xi)).

3. Administrative and Teaching Action Necessary to Achieve Objectives

A. Phase 1 - Implementation: Designer/Teacher

Although the same person may be functioning in different roles as curriculum designer, examiner and teacher it is easier to separate these roles for convenience and conceptual purposes.

There may be several intermediary links between curriculum designer and teacher, depending on the particular situation, type and size of the educational system but, direct or indirect, interactions between designer and teacher are most important.

Having stated the aims and objectives of the chemistry programme the implementation will depend on three principal factors - demographic, economic and social. Demographic factors will include, among others, numbers of schools and teachers, their rates of increase and pattern of distribution. Economic factors will include the national financial commitment to education, priority needs of scientific manpower at all levels, availability of additional resources from outside the country and supply of science equipment. Social factors will include societal attitudes to chemistry and its applications, readiness of the teaching profession for change, and patterns of authority and leadership at different levels.

The phasing and rate of implementation of change will be dependent on these factors. But the successful implementation of any programme will also depend on the effectiveness of communication between designer and teacher which will need to be clear, unambiguous and appropriate to the level and background of those concerned. Channels of communication should be established, if not already in existence, and developed to comprise a set of clearly specified operations. Since no communication system will function with complete efficiency it will be necessary to provice assistance and guidance in interpretation for teachers.

Concurrent with the establishment of effective communication between designer and teacher there should be administrative action to allocate resources for implementation. Among these will be additional teaching staff, in-service training facilities, guide lines for teachers, textual materials, specifications of examinations and above all, time. Only when this management action has been taken can the second phase of implementation - teacher/pupil change - be initiated.

B. Phase 2 - Implementation: Teacher/Pupil

The teacher becomes aware of the aims, objectives and content of a new chemistry programme through syllabuses, teachers' guides and schemes of work, pre- and in-service training courses, examination papers and work schemes, and textual or other resource material. It is assumed that the teacher is reasonably competent in the subject matter and in the normal way he would prepare lesson plans designed to implement the objective and content of the new course. It is also assumed that he has appropriate working space and facilities for teaching chemistry.

In presenting his material the teacher uses oral communication, textual materials, demonstrations, visual aids, pupil investigation, discussions, class and home exercises, extra curricula activities (such as science clubs, projects and visits to outside laboratories and factories) and examinations. Examinations comprise question/answer technique in the ordinary lesson, periodic tests and any final examination. These all provide learning situations and may be used to diagnose pupils' difficulties and to modify teaching methods. Moreover they may be used to assess how effectively the objectives have been achieved.

Implementation in Ceylon

Implementation of new chemistry curricula in the Ceylon system has been effected by detailed schemes of work for teachers. These not only listed the usual items in the syllabus but gave teaching outcomes for each section, suggested teaching procedures, experiments and activities for pupils and also highlighted problems of motivation, both for pupil and teacher, at appropriate points. Another interesting aspect is the home activities which pupils are encouraged to undertake - through the provision of simple apparatus and chemicals where necessary - not only to familiarise them with situations in everyday life in which chemistry plays a part, but also to assist direct transfer of classroom learning to everyday life.

Assistance and guidance to teachers are given through weekend study circles and vacation courses. Supervision of such assistance is provided by the service supervisory staff in consultation with the curriculum design staff of the Ministry of Education.

Further details of the administrative and teaching action carried out in Ceylon to achieve objectives of the chemistry programme are given in Part II - Supporting Evidence.

4. Planning an Examination Scheme

A. Introduction

Before an examination scheme is planned it is essential to know the purpose of the examination. Various ways in which examinations in chemistry are being used are given in Chapter 2 but each country will have to decide within its own educational system which of these purposes are intended and this information must be submitted to the examiners.

It is also important to have information on the characteristics of the student population for which the examination is intended. These characteristics will include age groups of students, length of schooling and other relevant socio-economic and educational factors. They will also include the pattern of previous assessments based on earlier examinations and evaluation procedures.

B. Information required by the examiner

Effective examining always used to be and often still is the inspired and intuitive work of successful teachers. However, in recent years, concepts and techniques have emerged which permit professional educationists to work on curriculum design and evaluation, exposing some of the weaknesses of earlier methods of examining. This has enabled the diverse experiences of professional curriculum designers and teacher examiners to reinforce each other leading to new concepts of examination construction and assessment.

In order to design an examination it is essential for the examiner to have some basic information from the instructional staff or from those who have designed the curriculum. In addition to the broad purposes of the examination and the characteristics of the student population (mentioned in Section A above) the examiner must know the content of the curriculum and the pupil activities involved.

One way of presenting course content is by means of a written syllabus organised so as to indicate major topics and sub-topics. There should be some clear indication in the syllabus on the importance and emphasis, as well as the time that has been devoted to the major segments of this course. The content specification should indicate the basic material required of all pupils and the alternate or optional material in the course.

In addition to specifications of content, it will be necessary to indicate the levels of attainment expected of the pupils. In particular, it will be necessary to specify in respect of each area of the content the level of behavioural achievement expected (cognitive, psycho-motor and affective factors are the usually accepted technical terms in this context (see glossary).

Furthermore, relevant and pertinent information can be given by describing the teaching procedure and activities to which the pupils have been exposed to achieve the objectives.

It is a growing practice in the planning of an examination scheme to relate the content objectives, behavioural objectives and teaching activities, in a variety of ways. The most common way by which this specification is achieved is by means of a two-dimensional grid, examples of which are given in Part II - Supporting Evidence, 1, 3 and 4.

C. Available forms of assessment

There are three broad categories of assessment:

(i) Internal assessment by teachers, which is an integral part of teaching.

(ii) Assessment by an external body. This is essentially a terminal process to assess attainment at the end of a course.

(iii) Assessment in which a terminal assessment administered by an external body is combined with an assessment made by the teacher.

The techniques of assessment in all these categories may be a written examination, an oral test, a practical examination based on written evidence or direct observation of pupils' activities. Various combinations of these techniques are possible.

D. The appropriate use of various forms of examination (see also Part III - Background Papers, 6)

(i) Written examinations differ widely in form. One type of question is called fixed response, that is a question in which the candidate has to choose between some determined responses, e.g.:

When a weighed piece of magnesium is completely burned in air, which one of the following is obtained?

(i) A black powder whose weight is greater than that of magnesium.

(ii) A black powder whose weight is less than that of the magnesium.

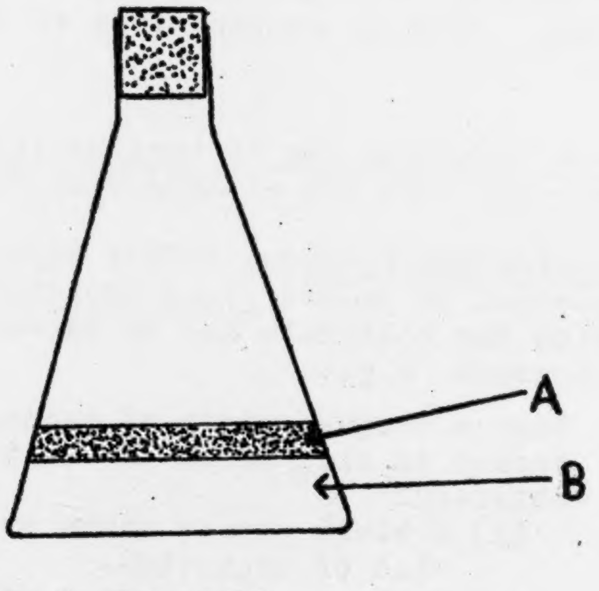
(iii) A white powder whose weight is greater than that of the magnesium.
(iv) A white powder whose weight is less than that of the magnesium.

A second type of question is the single open-ended question to which candidates have to give a short answer having been determined by the examiner and included in the mark scheme, e.g.:

> Name two substances in the Ceylon home that are used as fuels. Give one instance where a liquid fuel is used as a solvent in the home. State whether the solute you have mentioned in your example has undergone a physical or a chemical change.
>
> Give one reason to support your conclusion.

A third type may be called structured questions. In these the examiner has predetermined in detail the structure of the answers he expects and the question is divided into parts that correspond to these answers. In other words a single complex situation is broken down into a set of open-ended questions all requiring short answers, e.g.:

> In an experiment 30 ml of chloroform is shaken with 30 ml of NH_3/H_2O of strengths (a) 1M. (b) .9M. (c) .8m (d) .7M. (e) .6M. in a conical flask as shown in the diagram.
>
> The mixtures were allowed to stand for 5 minutes and analysed to find out the amounts of NH_3 in the two layers.



> (1) What liquid is in layer A?
>
> What liquid is in layer B?

(2) How would you obtain 10 ml of layer B without using a pipette? You should draw a diagram if it will make your answer clearer.

(3) How would you show that layer B contained ammonia?

(4) What might happen if a solution of ammonia in chloroform was shaken with pure water?

(5) What might happen if a solution of ammonia in water was shaken with a solution of ammonia in chloroform?

Mark Scheme

(1) A - water (1)
B - chloroform (1)
(2) Use burette (2)
Technique (3)
(3) Any moist (1) indicator (1) showing colour change (1)
(4) Movement of ammonia from chloroform into water (2)
(5) Movement of ammonia from chloroform to water (1)
Water ⟶ chloroform (1)
Possibility of equilibrium (2)

Total - 15

A fourth category of question is commonly called essay question. These questions usually require a fairly long answer, not necessarily in the form of a prose essay since other forms of communication - mathematics, diagrams or notes - may be more appropriate. This category may be subdivided according to the method of marking. If the answers are marked by a finely structured mark scheme largely predetermined by the examiner the amount of freedom of response is less than if the answer is marked by impression with a more general mark scheme when the candidates have opportunities of submitting more divergent answers. E.g.:

> A mixture of hydrogen and nitrogen, left in a jar for several days, at room temperature, shows no appreciable amount of ammonia. What special reaction conditions are used in the manufacture of ammonia from hydrogen and nitrogen, to make the production economical? Give one advantage of making the mixture of gases flow through the reaction chamber, rather than remain static.

It is obvious that if an examiner wishes a candidate to answer in a predetermined way, the fixed response and structured questions are the most useful. If, however, the examiner wishes to allow the candidate more freedom in answering, the free response questions are more appropriate.

It should be noted that fixed response tests lend themselves more readily to fine specification of most of the objectives, to pretesting and data processing techniques, particularly that of question analysis, and the quicker publication of results.

They allow a wide coverage of both content and objectives and the marking of such tests is highly reliable and requires little skill. Examining bodies, however, should think carefully about the possible effects of adopting this form of assessment from motives of administrative convenience alone, because there are many educational qualities that fixed response tests are unlikely to measure.

The "Essay" type questions, which are marked from a detailed marking scheme, require a special comment. While apparently allowing free response to the candidate, in practice the answer, if it is to be successful, must fit a detailed response largely predetermined by the examiner. It may seem that if an examiner has already structured the question in his own mind it would be better to make this structure explicit to the candidates in the question rather than to ask the candidates to guess at the structure.

(ii) Oral tests

In some countries much use is made of oral techniques as part of a total assessment of achievement. This has the advantage, not only of testing knowledge and understanding, but also of allowing direct feedback from examiner to candidates during the examination, so that it closely resembles a teaching-learning situation, particularly in chemistry, if it is associated with a practical task.

(iii) Assessment of practical work

The assessment of skills, abilities and attitudes in practical work requires special mention. The marking of written evidence after a practical exercise has obvious defects as a means of assessing the practical work of the candidate. Examiners may wish to consider the use of a direct observation technique for this purpose in the form of a continuous assessment of practical work by teachers. It should be pointed out that this assessment involves all three domains, cognitive and affective as well as the psychomotor behaviour.

There is some evidence to support the view that a pupil uses abilities when he is carrying out experimental work which are, to some extent, distinct from those used in non-practical situations. (A suggested specification of these objectives is given in Part II - Supporting Evidence, 4, together with some information on the measurement of practical ability by continuous assessment, and a brief discussion of the problem of moderating these assessments.)

If, however, only the psychomotor dimension is to be evaluated, there is much experience in industry where the evaluation of skills is extensive.

E. The effect of different forms of assessment on the teaching/learning situation

The emphasis placed on the different techniques of assessment differs between internal and external assessment. Assessment within a school relies heavily on oral and other direct observation techniques, while in external assessment the emphasis is on written examinations and tests.

It is also apparent that in external examinations there is a general trend at present towards a more widespread use of fixed response questions in objective tests (so called). This trend towards a more rigid candidate-response in examinations is in contrast to recent curriculum reform which gives greater freedom of response and initiative to the pupils in their lessons. Taken to their extreme, these two trends are not compatible, and the dangers in this situation can only be countered by the use of more balanced forms of final assessment in which more than one technique of assessment is used.

If examinations are to have the desired backwash effect on education, the techniques used in examining must resemble those used in teaching; so that if a teacher finds it necessary to use more than one form of questioning the examiner should do likewise. The use by an examiner of one technique alone may well lead to coaching in schools in this technique to the exclusion of other, and equally desirable, forms of questioning between teacher and pupil.

One obvious example of the effect of a form of assessment on the teaching/learning situation is that of direct observation of the pupil over a period of time. If this is carried out by the teacher, and if the outcome is used as a part of the final assessment of the pupil, a change in the relationship between pupil and teacher is almost inevitable. A pupil who knows that his teacher is also his examiner is bound to view his teacher in a different light. This change in relationship may not be undesirable, but it should not be ignored by those who wish to introduce some form of continuous assessment which will subsequently be used by external authorities.

5. Administration and Execution of Examinations in Chemistry

A. The appointment of appropriate individuals to examination boards

If the examination is truly to be an instrument of wise education, if it is to serve as a means of assessment of educational achievement and as a means of making adjustments to the educational process, then the composition and qualifications of examiners and examination boards must be carefully considered.

An examiner or a moderation board member should possess, not only long experience in teaching and high competence in his subject field, but, more importantly, he should feel a strong moral obligation of being open to new ideas and to innovations in the educational process. He should consider the examination not only as an instrument of performance but as a means for educational reform.

To sustain among examiners these positive attitudes of openness and dedication to improvement, they should periodically study their techniques in the light of the wide international ferment in the field of curriculum development and assessment. In keeping with this self-critical spirit, several practical steps can be recommended to examination boards:-

(i) They should establish liaison with members of the scientific and technological communities that education professes to serve (for example, with doctors and health officers, with engineers, with research scientists, with industrial managers, etc.).

(ii) They should maintain working contact with university scientists and with science curriculum design groups where these exist.

(iii) They should occasionally consult an international expert on examination construction and use.

(iv) They should keep abreast of the growing specialised literature on assessment and examinations in science.

The reasons for stressing these points and recommending these steps are that in some countries - indeed in many countries - the Examining Boards have cut themselves off from such stimulation. As a result they tend to become a positive blockage to innovation and a source of discouragement and frustration to those who are striving for curriculum development. This can be particularly acute in the context of a rapidly developing scientific discipline such as chemistry.

B. Sequence of operations

Having designed a plan and an organization, the examiner is ready to construct and administer the examination. The general sequence of operations and the various forms which each operation may take are described below.

The sequence of operations may be common to all countries, but the form of each operation is likely to vary from country to country, depending upon their aims, resources and experience in examining.

Figure 2 is an attempt to show the sequence and possible forms of operation diagrammatically. The sequence of operation is shown vertically and, for the sake of completeness, a construction of the specification is included, although this has already been discussed in a previous section.

The various forms of operations available to an examination organisation are shown horizontally. Those who are about to set up their own examinations for the first time will need to choose which form of operation is most appropriate to them; for convenient usage the forms of operation are arranged from left to right in order of increasing difficulty. By their very nature, the order reflects, in part, the historical development of examination construction. It should be noted, however, that a country need not evolve its examination structure by passing progressively through the sequence shown in Figure 2. Rather it should choose that form of operation which is most appropriate to the present circumstances.

To illustrate this point the development of operations in chemistry examining in Ceylon over the past eight years is shown in Figure 3. The reader may find it helpful to draw a similar diagram for the examination system in which he is interested.

C. Writing the questions

The first job of an examiner is to arrange for some questions to be written and the simplest way is for one or two chief examiners to write them. For short answer and free response

FIGURE 2

Sequence of Operations

showing sequence (vertically) and form (horizontally) of operations available to an examination organisation

1. Specification (content)	Syllabus	Syllabus weighted	Themes	Teaching materials	Detailed survey of field practices
1.1 Specification (objectives)	Cognitive	Psychomotor	Affective		
2. Writing questions	1 or 2 examiners	Writing group	More than one writing group	Central pool of questions	International pool of questions
3. Organisation of critical review of questions	Moderation by subjective judgment of a small group	Organised shredding by groups of writers and moderation	Pre-Testing	Moderation with Pre-Test data	Shredding with Pre-Test data
4. General administration of examinations	School based.	Regionally based	Nationally based	Multi-lingual national examination	
4.1 Security.	Not secure	Secure before examination	Secure before and after examination		
4.2 Printing and handling	Free response	Structured questions	Fixed response questions	Complex examination form	
4.3 Marking	Fixed response	Structured questions	Free response	Complex examination form	
5. Analysis/ interpretation/ reporting results	Pass/Fail using a single total mark	Grading using a single total mark	Profile of a candidate in terms of stated objectives in chemistry		

FIGURE 3

The development in Ceylon of operations outlined in Figure 2

1960

1	▨					
1.1	▨					
2	▨					
3	▨					
4				▨		
4.1		▨				
4.2	▨					
4.3			▨			
5		▨				

1962

1	▨					
1.1	▨					
2	▨					
3	▨					
4				▨		
4.1		▨				
4.2	▨					
4.3			▨			
5		▨				

1965

1	▨	▨	▨	▨		
1.1	▨					
2	▨					
3	▨					
4				▨		
4.1			▨			
4.2	▨		▨			
4.3	▨		▨			
5		▨				

1970

1	▨	▨	▨	▨		
1.1	▨	▨				
2		▨				
3	▨	▨				
4				▨		
4.1			▨			
4.2	▨	▨	▨			
4.3	▨	▨	▨			
5		▨				

questions this method may work quite well, although there is always a danger that the examination will reflect personal likes and dislikes of the examiner.

If it is proposed to use fixed response and structured questions, however, examining authorities may find that the use of one or two question writers will be effective for only a short time. The construction of such questions requires skill, experience and, above all, plenty of time. And if feed-back from classroom to examination is to be attained, the use of one or more groups of question writers, directly in touch with teaching in schools becomes desirable.

The transition from the use of one or two question writers to a group of question writers is at present taking place in Ceylon, where the first step to bring about this change has been the training of question writers. This was the aim of an evaluation workshop conducted in Ceylon in August 1968, a report of which is given in Part III - Working Papers, 6. This workshop is not presented as a model, but rather as an example of one attempt at the sort of training which is necessary.

The selection of question writers is an important part of the process of building up an examination. It should not be taken for granted that because a person is a good chemist, or a good teacher, or both, that he will necessarily be a good writer of questions for a chemistry examination. It is only by judging efforts in the actual construction of questions that good question writers can be found; and, if it is thought that examinations should reflect and encourage the desired objectives and approach to education in chemistry, then the more the writers are in tune with actual teaching situations the better.

It cannot be denied that the writing of fixed response and structured questions is difficult and time consuming, and writers before long may run out of ideas. It may appear, therefore, that the expenditure of good questions year by year, never to be used again, is wasteful. For this reason it is a growing practice to collect a pool, or bank, of questions, properly classified, which can be used when required, and possibly used more than once. To extend this idea, there seems no reason why an international pool of questions at various levels of chemistry should not be formed (it should be noted that pre-testing, if required, would have to be carried out again on an appropriate population). An international pool of questions could greatly accelerate the growth of examination construction in developing countries. It must be pointed out, however, that the formation of a pool of questions requires more complex systems of filing and security than does other forms of examining, and this should be borne in mind when considering the administrative requirements of a new examination system.

D. Critical review of questions

The writing of questions is only the first part in the construction of an examination. The next is a critical review of the questions, a stage which is particularly important for questions of the fixed response and structured type in which much chemistry is often contained in the question itself. If question writers are inexperienced, there will be a large wastage of questions at the review stage, often as high as three questions out of four.

The first critical review of questions is often undertaken by a small group of examiners, but the practice varies from country to country. A possible set of criteria for a first review has been worked out for use in Ceylon examinations in the Ceylon Workshop and is set out below. (The list is not intended to be comprehensive and other factors may have to be considered, particularly some technical factors in the construction of fixed response questions. If necessary, the reader should refer to standard publications on test construction - see bibliography.)

(i) Accuracy of the chemistry.
(ii) Simple errors of construction, e.g. language, ambiguity, obvious clues, etc.
(iii) The appropriateness of the level of study.
(iv) The time required to read the questions.
(v) The specification of the question.
(vi) Its effect on the teaching/learning situation.

E. Construction of the first draft of examination paper

The reviewing process (sometimes called shredding) may be repeated by other similar groups of people. Then the examiner is in a position to construct the first draft of the examination paper.

In doing so, the major factors, in addition to those given above, which must be kept in mind are:-

(i) The specifications.
(ii) The range and distribution of difficulties of the questions.
(iii) Duplication of questions.
(iv) Overall time required to complete the paper.
(v) The relevant weighting in terms of marks for the various kinds of questions and the weighting within each question.
(vi) Alternative questions and questions from which a choice can be made should be equivalent as far as objectives, time required and difficulty are concerned.

F. Moderation of papers

Another review of questions and of the text as a whole, similar to that conducted by the examiner, may now be undertaken by another group, normally called moderators, before the publication of the examination in its final form.

For this meeting of the moderators the following materials would be required:

(i) The draft examination.
(ii) The table of specifications.
(iii) The marking schemes.

(It is assumed that the background material available to the examiner is also available to the moderators.)

G. Pre-testing of questions

The sections D, E, F above are based on the assumption that pre-testing of questions is not feasible. However, in some educational systems it is possible to carry out pre-testing of questions, particularly of the fixed response type.

Pre-testing is essentially a trial of questions on a representative sample of pupils, the objective being to provide the examiner with an additional tool in the form of quantitative data, which will assist him to refine his examination.

For details of pre-testing, readers should refer to the literature, but they may find of interest the procedure adopted for the first time in Ceylon in 1968 as a preparation of the Evaluation Workshop held in August of that year (see Part III - Background Papers, 6).

Among the main data concerning an individual fixed response question which arise in pre-testing are:

(i) Percentage response to each sub-section (distractor) and a difficulty index (D.I.) of the correct answer.
(ii) The discrimination coefficient (D.C.).

The difficulty index tells the examiner the percentage of pupils who responded to a question in the pre-test in the way the examiner expected. A low difficulty index means few questions answered correctly.

The discrimination coefficient tells the examiner the correlation between the performance of candidates in a particular question with the performance of candidates in the test as a whole. Neither the difficulty index nor the discrimination coefficient tells the examiner that the response was educationally desirable nor does it tell him that the chemistry in question was sound. Thus, it will be found in the examples given below that a pre-tested question whose difficulty index and the discrimination coefficient were quite satisfactory had to be rejected on other grounds.

With the aid of this additional data moderators would then proceed in the usual way. (It is assumed of course that both examiners and moderators are competent in interpreting pre-test data.)

H. Critical review of some fixed response questions from the Evaluation Seminar held in Ceylon in 1968

Questions

1. Sulphur powder and iron powder were strongly heated together. When tested for iron and sulphur, their properties were NOT evident. This is because:
   - (i) a mixture was formed.
   - (ii) a compound was formed.
   - (iii) an element was formed.
   - (iv) a physical change had occurred.

2. An element reacted with cold water displacing hydrogen. This element is most likely to be:
   - (i) high up in the activity series.
   - (ii) just above hydrogen in the activity series.
   - (iii) just below hydrogen in the activity series.
   - (iv) very far down in the activity series.

3. Whenever a liquid becomes a solid which one of the following can be stated as being certain to occur?
   - (i) Arrangement of particles in a regular pattern.
   - (ii) Packing of particles into a smaller volume.
   - (iii) Absorption of thermal energy from the surroundings.
   - (iv) Decrease in the kinetic energy of the particles.

4. A gas showed the following properties. It immediately extinguished a glowing splint. It had no reaction with limewater. When metallic copper was heated in the gas, no visible change was observed. Which one of the following could the gas be?
   - (i) Sulphur Dioxide.
   - (ii) Nitrogen.
   - (iii) Carbon Dioxide.
   - (iv) Oxygen.

5. $NaOH + H_2SO_4 = NaHSO_4 + H_2O + X$ cal.

$NaHSO_4 + NaOH = Na_2SO_4 + H_2O + Y$ cal.

$2\ NaOH + H_2SO_4 = Na_2SO_4 + 2H_2O + Z$ cal.

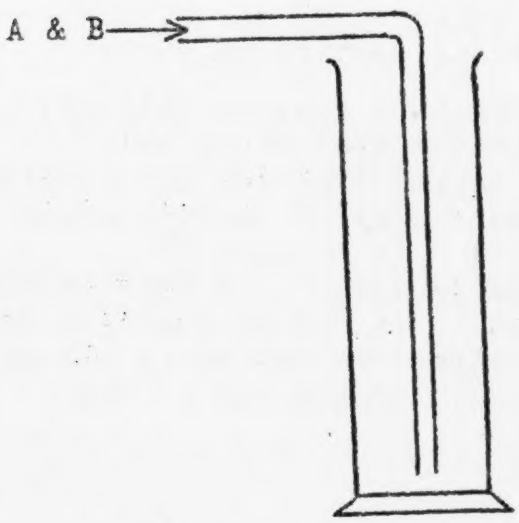
The above equations indicate 2 methods by which sodium sulphate may be prepared. Which one of the following describes the thermal energy relations of the two processes?

(i) X - Z = Y.
(ii) Y + Z = X.
(iii) Y - X = Z.
(iv) X + Y = Z.

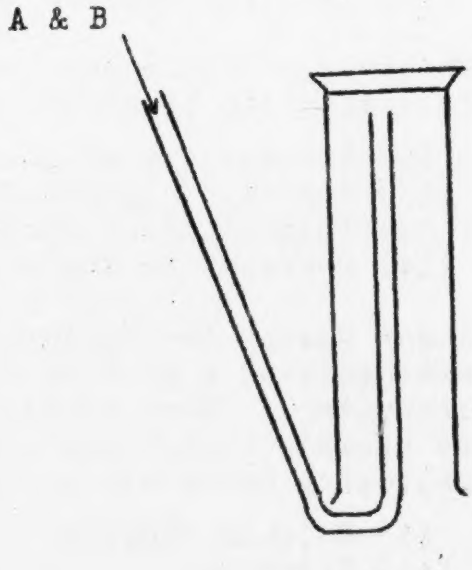
6. Gases A and B have the following physical properties.

	Solubility in water	Boiling point	Density (relative to air)
Gas A	70 g per litre	-110°C	0.8
Gas B	15 g per litre	2°C	1.2

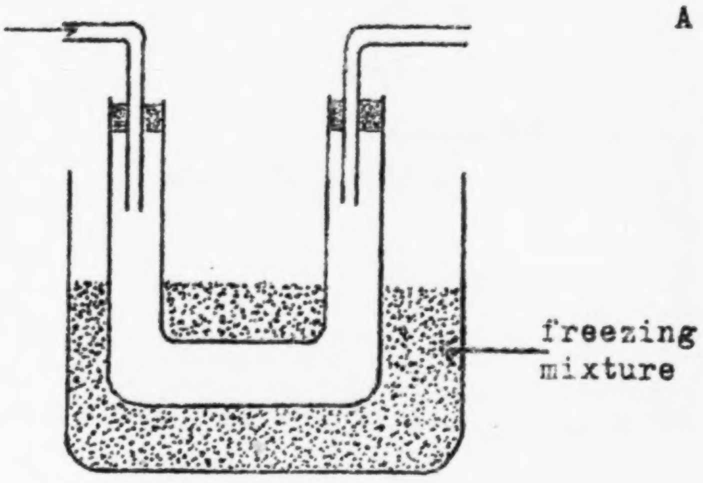
In order to collect as pure a sample of B from a mixture of A and B with as little wastage of it as possible, which one of the following collection procedures would you adopt?



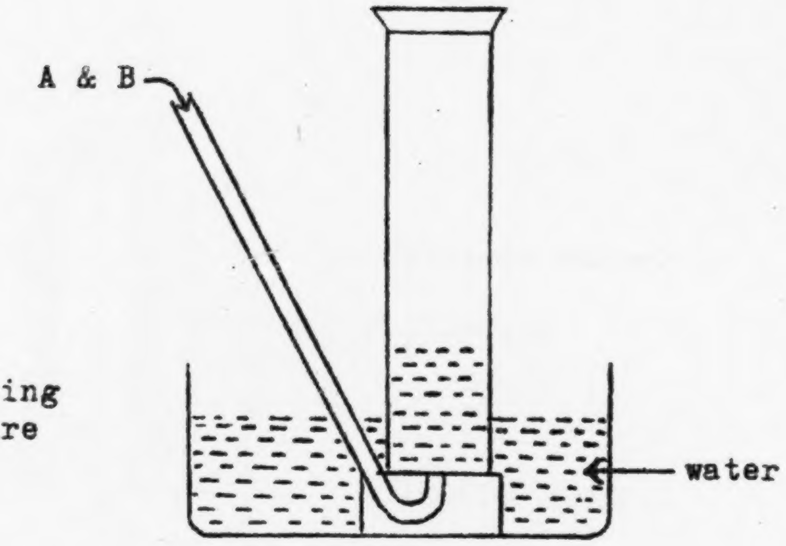
(i)



(ii)



(iii)



(iv)

7. What weight of potassium chloride (molecular weight 74.5) would be formed when excess hydrochloric acid reacts with 17 g of potassium nitrite KNO_2 (molecular weight 85, the salt of the weak acid nitrous acid)?

   (i) 7.45 g
   (ii) 8.5 g
   (iii) 14.9 g
   (iv) 17 g

8. In which one of the following solutions will chlorine NOT be liberated?

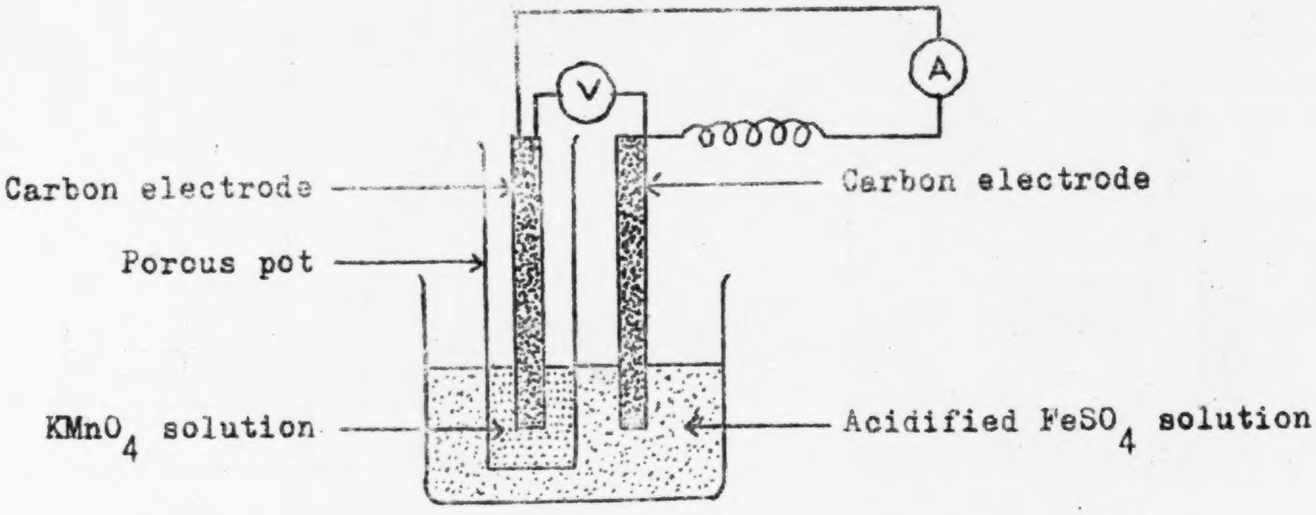
   (i) Cold concentrated hydrochloric acid and hot concentrated sulphuric acid.

   (ii) Cold concentrated hydrochloric acid and bleaching powder.

   (iii) Cold concentrated hydrochloric acid and crystals of potassium permanganate.

   (iv) Concentrated hydrochloric acid heated with manganese dioxide.

9. The diagram represents an electrical circuit involving an oxidation reduction cell C, a voltmeter V, an ammeter A and a resistance R.


Carbon electrode
Porous pot
KMnO4 solution
V
A
Carbon electrode
Acidified FeSO4 solution

In which of the following ways would variation of concentration of the permanganate solution be expected to affect the voltmeter and ammeter readings?

(i) The ammeter reading will change but the voltmeter reading will remain unchanged.

(ii) The voltmeter reading will change but the ammeter reading will remain unchanged.

(iii) Both the ammeter and the voltmeter readings will change.

(iv) Neither the voltmeter nor the ammeter readings will change.

10. When four substances A, B, C, D were heated in air the following observations were made. Which one of the following observation would alone suggest that the substance concerned was NOT an element?

(i) A liquified but no new substance was formed.
(ii) B burned producing sulpher dioxide.
(iii) C burned producing only an oxide.
(iv) D neither liquified nor produced a new substance.

Review

Item No.	Content code	Objective	Accuracy	Errors	Level	Time	D.I.*	D.C.
1	4	(1)	✓	-	✓	✓	.69	.54
2	14	(1)	✓	Needs editing	✓	✓	.71	.50
3	2	(2)	✓	-	✓	✓	-	-
4	29	(2)	✓	Stem may be shortened	✓	✓	.63	.39
5	27	(3)	✓	Construction (clarify)	✓	✓	.44	.39
6	2	(4)	✓	Obvious clues (i) & (ii)	✓	✓	.46	.44
7	18	(3)'	✓	Use formula weight instead of molecular weight. Irrelevant data.	✓	✓	-	-
8	29	Changed from (2) to (1)	Questionable	Obvious clues	✓	✓	.55	.40
9	19	(3)	✓	R omitted in diagram	Too difficult	Too long	.46	.28
10	4	(4)	✓	Needs editing	✓	✓	.34	.44

* Multiply by 100 to obtain percentage figures for difficulty index

Content Code:

TABLE OF SPECIFICATIONS - CONTENT CODE

1	Nature and scope of chemistry.
2	Substances and their properties.

3 Solutions.
4 Elements, mixtures and compounds.
5 Physical and chemical change.
6 Burning.
7 Air.
8 Oxygen
9 Conditions for combustion.
10 Oxidation - reduction I.
11 Action of acids on metals.
12 Hydrogen.
13 Laws of chemical combinations.
14 Action of water on metals.
15 Atomic theory.
16 Classification of reactions.
17 Chemical calculations.
18 Action of acids on salts.
19 Electrolysis.
20 Ionization and equilibrium.
21 Corrosion.
22 Atomic structure.
23 Carbon chemistry.
24 Chemicals from air.
25 Chemicals from sea.
26 Chemicals from soil.
27 Energy changes in chemical reactions.
28 Radiochemistry.
29 Oxidation - reduction II.

Objectives code:

(1) Recall.
(2) Comprehension.
(3) Application.
(3)' Calculation.
(4) Analysis, synthesis and evaluation.

Additional comments on the learning/teaching situation

Q.1. Question based on actual experimental work done in class. Encourages pupil activities/ teacher demonstrations, making observations and drawing conclusions from experimental data.

Q.2. May be answered without actually doing the experiment, but it encourages the teacher to introduce simple experimental work. The reference cold water suggests a further investigation of the effect of variables such as temperature. The item could be made more concise as follows:-

An element . . . . . . The location of the element in the activity series is most likely to be:

(i) high above hydrogen.
(ii) just above hydrogen.
(iii) just below hydrogen.
(iv) far below hydrogen.

Q.3. Brings up the idea of using a model in chemistry to explain the properties of matter. When generalisations are made, as suggested in this item, the teacher should be encouraged to be conscious of any anomalous situations such as the behaviour of water in this instance.

Q.4. The item could be more concisely stated. Encourages pupil activities/teacher demonstrations, make observations.

Q.5. Needs minor editing. The question indicates two methods, but gives three equations. It is necessary to indicate that the first two equations refer to one process.

This question draws attention to broad generalisations and applications.

Q.6. Emphasises the need to confront pupils with the ability to cope with new situations. Encourages training in suitable designing of experimental set-up to solve problems and apply general principles in chemistry to new situations. Unfortunately, choices (i) and (ii) are not good distractors.

Q.7. Editing:

(i) Delete - "the salt of the weak acid, nitrous acid" (irrelevant data).

(ii) Use formula weight instead of molecular weight since the substance is ionic.

(iii) Use 8.5 instead of 17. It will be easier for the pupil to recognize the relation between the formula weight of KNO_2 and the given weight.

Item encourages the teaching of chemical calculations.

Q.8. Example of an unsatisfactory question. (ii), (iii) and (iv) are obvious clues, actually carried out by pupils, but (i) is not done and on that basis alone it may be selected. Accuracy of the chemistry is also questionable. Any attempt to test (i) experimentally by a pupil may be dangerous.

Q.9. Question rejected. May be too difficult. The examiner probably wants to test the understanding of variation of voltage with change in concentration, which may be too difficult at this level. In addition, he has coupled to it a problem involving variation of both voltage and current which makes it even more difficult.

The diagram does not show 'R'.

Q.10. Editing: Insert the word 'separately' after 'heated'. A good question which calls for analysis synthesis and evaluation of experimental data and making conclusions.

I. Rejection of fixed response questions

Rejection of questions may be effected at a relatively unsophisticated level on the following grounds which are illustrated by examples.

(i) Accuracy of the chemistry

Copper reacts with 50% nitric acid according to the equation:

$$3Cu + 8HNO_3 = 3Cu\,(NO_3)_2 + 4H_2O + 2NO$$

The volume of nitric oxide at STP which is formed from 63.5 g of copper would be:-

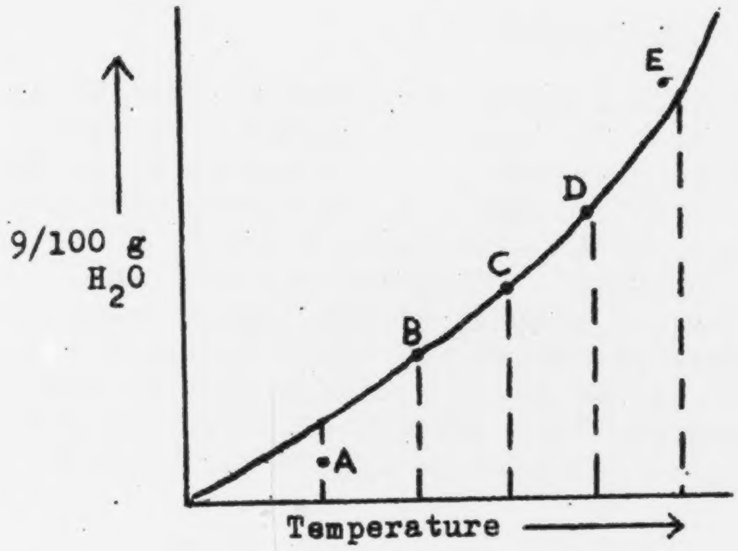
(i) 22.4 litres.
(ii) 44.8 litres.
(iii) 14.9 litres.
(iv) 7.5 litres.

Comment

This reaction is more complicated than indicated by the equation given in this item.

(ii) Simple errors such as language, ambiguity, obvious clues, etc.

The solubility of potassium chloride at various temperatures is represented below:-



Which one of the following points indicate a saturated solution?

(i) B and D only.
(ii) B, C, D and E.
(iii) C and D only.
(iv) A only.

Comment

Both diagram and the stem of the question are confusing. The correct response does not appear as one of the choice.

(iii) Appropriateness of the level of study

i	ii	iii	iv	v	vi	vii	viii	0
a	b	c	d	e				
f	g	h	i	j				
k	l	m	n	o				

Members of the first elements in the periodic table are represented above.

Which one of the following pairs of elements is LEAST likely to show resemblance in their chemical properties?

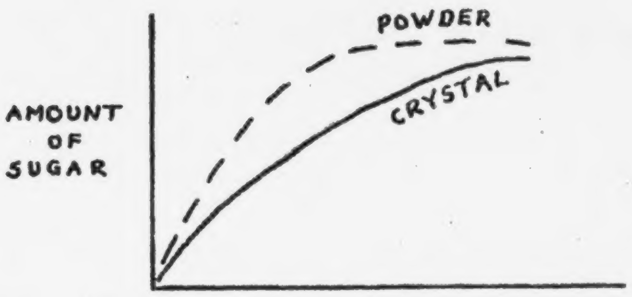
(i) f, k.
(ii) g, l.
(iii) f, l.
(iv) k, g.

Comment

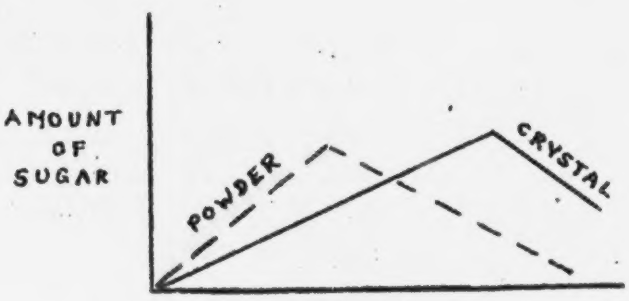
In the Ceylon teaching manual this depth of study is not expected to be reached.

(iv) Reading time of questions

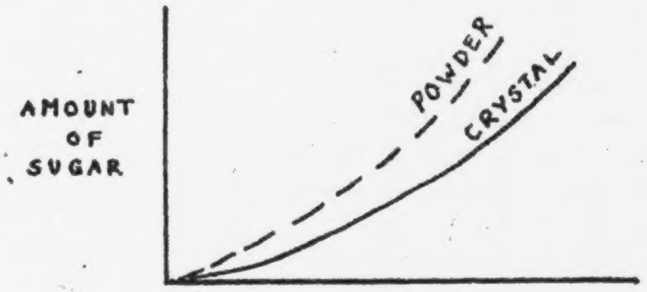
A pupil weighed 10 g sugar into each of two 100 ml beakers A and B. The sugar in beaker A was in the form of a fine powder. The sugar in B was in the form of a single lump. He simultaneously and carefully poured 80 ml water into A and B, without stirring the contents. At intervals of 5 min. duration, five-drop samples of the liquid from each beaker were withdrawn and their separate sugar content determined. If these results are graphically represented, which one of the following would best depict the results of the experiment?



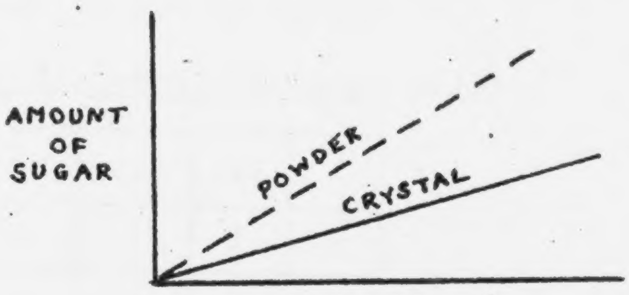
(i)



(ii)



(iii)



(iv)

Comment

An average pupil may take as much as 1½ minutes to read and understand the stem only of this question. In this paper 45 minutes are allocated for answering 40 questions and the time taken for this question is out of proportion. Furthermore, there is an air of unreality in the question; it seems unlikely that the writer of the question actually performed the experiment.

(v) Effect on the teaching/learning situation

When H_2S is passed into aqueous solutions of the following groups of metals which group will produce only black precipitates?

(i) Pb, Cu, Cd.
(ii) Cu, Co, Sb.
(iii) Pb, Ni, Cu.
(iv) Ni, Cu, Zn.

Comment

This kind of item encourages memorising information, which in the context of school chemistry can only be regarded as trivial.

At a more sophisticated level when pre-test statistical data are available further criteria may be used as exemplified below.

(i) Low difficulty index

When a finely divided metal A is mixed with the oxide of metal B and heated, no change is seen to occur. When the oxide of metal B is heated, with a metal C, metal B and oxide of C are formed. When metal C is added to a solution of a salt of A, particles of metal A are deposited. From these observations the increasing order of activity of the metals B, C and A is:-

(i) C, B, A.
(ii) B, C, A.
(iii) C, AB, B.
(iv) A, B, C.

Comment

This item though discriminating positively (0.23) has a very low difficulty index of 0.13. This is probably due to the length of the item, the complexity of information and the use of symbols.

(ii) Low discrimination coefficient

What may NOT always be expected in a balanced chemical equation describing a chemical reaction is that the sum of the:-

(i) weights of the products equals the sum of the weights of the reactants.

(ii) volumes of the products equals the sum of the volumes of the reactants.

(iii) number of atoms of the products equals the sum of number of atoms of the reactants.

(iv) atomic weights of the products equals the sum of the atomic weights of the reactants.

Comment

This item has discrimination coefficient as low as -0.33 although the difficulty index is 35%. Here the stem in the negative form may have been the significant contributory factor.

(iii) Non functioning distractors

Three similar burning tapers were plunged simultaneously into 3 separate cans containing coconut oil, petrol and kerosene respectively. Petrol caught fire at once, but NOT the other two. This was probably due to the fact that:-

(i) the temperatures of the three tapers were not the same.

(ii) the ignition temperature of petrol is higher than that of the other two.

(iii) the ignition temperature of petrol is lower than that of the other two.

(iv) petrol has a lower density than either of the other two.

Comment

The first response does not appear to function adequately as only 0.6% of the test population selected this choice. The fact that the first choice is a wrong response may be self evident from the first few words of the stem.

(iv) Over functioning distractors

Milton is used to remove ink stains. The agent responsible for the bleaching is active:-

(i) chlorine.
(ii) oxygen.
(iii) hydrogen.
(iv) hydrochloric acid.

Comment

72% of the test population selected the first choice (an incorrect choice) whereas only 20% responded to the correct choice. This type of over functioning is sometimes due to the poor wording of the item, as borne out here.

SPECIAL NOTE

Although an item may appear acceptable on indices of pretest data, it may still be possible that the item is unsuitable because of errors which escaped a first scrutiny. This could happen because the target population ignored the errors and responded correctly. But still the question must be either rejected or re-written even though the item appears to function well.

e.g. $NaOH + H_2SO_4 = NaHSO_4 + H_2O + X$ cal.
$NaHSO_4 + H_2SO_4 = Na_2SO_4 + H_2O + Y$ cal.
$2\ NaOH + H_2SO_4 = Na_2SO_4 + H_2O + Z$ cal.

The above equations indicate two methods by which sodium sulphate may be prepared. Which one of the following describes the thermal energy relations of the two processer?

(i) $X - Z = Y$.
(ii) $Y + Z = X$.
(iii) $X + Y = Z$.
(iv) $Y - X = Z$.

J. Administration of examinations

Organisations for the administration of examinations are bound to vary greatly. For the purpose of this report only those aspects of administration which may directly affect decisions on the form of examination are considered.

A general pattern of examination administration may be classified as school based, regionally based, and nationally based. If the assessment includes some form of continuous assessment it will entail an extra administrative load on the schools.

A particular difficulty in some countries may be the need to translate an examination into several languages. This produces problems not only in translation of idiom but also in translation of technical vocabulary. Other problems in question setting arise when there are different cultures within a country, particularly when referring to chemicals which may be found in one cultural background but not in another.

Security

Most examinations at present remain secure before administration. There are exceptions, the most important of which are some oral and practical examinations and continuous assessment of practical and other work which is the normal part of a chemistry course.

In some secure examinations, pupils may be allowed foreknowledge of a part of the examination or relevant materials of the examination; for example, open-book examinations and examinations in which books on data are allowed. At the other extreme some examinations, particularly some of the fixed response type, are kept secure before and after the examination. This raises further administrative difficulties.

Printing and handling of examinations

The degree of administrative difficulties in this area depends largely on the form of the examination and question used. For example, the use of fixed response and structured questions, giving a lot of various types of information in the question itself, inevitably leads to a bulky and more complex examination. The sheer bulk raises administrative problems in printing, packaging, storage, despatch and security.

Marking

Marking may be in three forms - from a fixed key, from a finely structured mark scheme and by impression, these being in the order of increasing difficulty.

A special comment is required on impression marking. This is a grading process rather than a marking process and is normally conducted by more than one examiner with general, rather than detailed, suggestions for marking. Such marking may call for a close examination of standardization procedures. People qualified to mark by impression are not always easy to find. A further administrative difficulty in impression marking arises from the need to post scripts from one marker to another in a short period of time.

Where the same question appears in more than one language in an examination paper and there is error in one translation, the examiner must classify all translated questions as faulty, because bilingual students may have read faulty and/or the correct one.

K. Analysis, interpretation and reporting of results

After marking, the examiner has available for analysis the marks of each candidate in each question, the total marks in each section, and the total mark in the examination as a whole. The common practice is to use the total marks scored as a measure of attainment. This implies that marks obtained on different sections of the test which were designed to measure different objectives can be added to give a meaningful result. The candidates are then put in order of merit according to the total marks. The simplest interpretation of this order of merit is to classify candidates into those who have passed and those who have failed. But interpretation may be used for many different purposes (see Chapter 2) and one may ask if it is the proper function of the examiner to decide who has passed or failed.

A second and more refined form of analysis puts candidates into grades indicating more clearly a hierarchy of performance by the candidates in the examination as a whole, but gives no further information on the candidates' performance in the different objectives of chemical education, in other words a profile of the candidate. To do the latter would require a more elaborate but possibly more profitable form of analysis, interpretation and reporting of the results, in the light of the objectives of chemistry teaching.

6. Strategy and Tactics for Curriculum Reform

The major assumption underlying recommendations and suggestions in this report is that curriculum development and examination reform are no longer one-stage operations. The pace and pattern of change is such that institutional arrangements which have been discriminately designed to cope with change have become imperative. One feature which should characterize such institutional developments is the provision for enhanced participation by teachers of chemistry, as well as professional chemists from universities, health, agriculture and industry.

Just as in chemical research, team work is increasingly replacing individual isolated efforts, so in curriculum development and examinations reform the need to provide for co-ordinated group effort is becoming necessary. This is being achieved through the establishment of curriculum development centres in which participants from many fields pool their experience and insights in the preparation of curriculum materials.

One of the prime functions of such activity may be the reduction of the 'lag time' between the emergence of new ideas and techniques, and their application in classroom situations. Another important function of curriculum reform groups is to sustain the momentum created during the process of renewing curricula, so that curriculum development groups might undertake to keep their materials, practices, methods and techniques in a process of continual reappraisal and readjustment.

A danger to be avoided is the elevation of the 'new' material and practices to the status of a new dogma. Hence, the need exists for some sensitive feed-back of information as to the direction in which the system is moving and a discriminating examination design provides such a monitoring and steering device.

Institutional organization for curriculum reform and examination design will not in itself ensure the diffusion and adoption of the new ideas. There is a need for more understanding of the process by which change comes about in school chemistry teaching. Inevitable consequences of curriculum reform are that relationships between all participants in education will need to be modified and adapted to permit growth and development.

APPENDIX A

GLOSSARY

Achievement test:	A test designed to measure the performance of a pupil in relation to the expected outcomes of instruction.
Affective behaviour:	A pupil's emotional feeling tone. This may include internal feelings and emotions, or may be expressed in overt behaviour marked by an immediate impulse to act.
Attainment test:	See Achievement test.
Behavioural objectives:	The analysis of the objectives of a course in terms of intended changes in the ways in which the pupil will feel and act.
Cognition:	An intellectual process whereby a pupil becomes aware and obtains all the various modes of knowing, including all the forms of reasoning.
Content objectives:	The listing of the major divisions and sub-divisions of the subject matter to be covered in a course. These are often included or implied in a syllabus.
Diagnostic test:	A test used to discover areas of strength and weakness in the attainment of the expected outcomes of instruction. Synonymous: Placement test.
Difficulty index:	See Item analysis.
Discrimination coefficient:	See Item analysis.
Facility index:	See Item analysis
Grid:	See Table of specifications.
Item:	A question, an incomplete sentence or other task in assessment designed to elicit a response from the examinee.

Item analysis:	A process by which the results of an examination are analysed for the purpose of identifying suitable items in terms of their difficulty and discriminating power. Difficulty index is usually expressed as a percentage of the population answering the item satisfactorily. It is also called facility value. The discriminating power (or coefficient) of an item is a measure of how well the item helps to differentiate between "good" and "poor" pupils, "good" and "poor" being determined on the basis of performance on the entire test.
Moderation:	A process in which items in a draft of a test or a grade obtained by assessment are critically scrutilized by a group of competent persons to ensure that all the specifications demanded are met.
'O' level:	General Certificate of Education at Ordinary Level - an examination taken at the end of a five-year secondary school course (aged 16 years usually).
Objectives:	The goals or aims of instruction or teaching stated explicitly (or sometimes only implicitly) in functional or behavioural terms. There may be several kinds, of which content objectives and behavioural objectives are the most widely used.
Placement test:	See Diagnostic test.
Pre-test:	A process of administering a try-out form of a test for the purpose of obtaining experimental data to improve the final form.
Psychomotor activity:	Co-ordination of a mental process and a motor activity.
Raw score:	An empirical score obtained by direct application of a marking scheme or key.
Scheme of studies:	An expanded syllabus outlining the contents and indicating the relative emphasis of time to be devoted to each division or sub-division of the subject matter.

Shredding:	The use of certain criteria to analyse the function of a question with a view to re-writing the question in a most suitable form.
Syllabus:	A document compiled by a curriculum authority, specifying the major divisions and sub-divisions of the subject matter to be covered in the course.
Table of specifications:	A diagrammatic analysis of the content and behavioural objects. Synonymous: Two-dimensional analysis, Grid.
Teachers' manual:	A scheme of studies explicit to include helpful methods and techniques of instructions, teaching aids and references to source materials. Synonymous: Syllabus of instructions.
Two-dimensional analyses:	See Table of specifications.

APPENDIX B

BIBLIOGRAPHY

Alles, J. et al. - Theoretical Constructs in Curriculum Development and Evaluation; Ministry of Education, Colombo, Ceylon, 1967, 24 pp.

Alles, J. - Notes on Structural and Functional Aspects of an Educational System relevant to Educational Administration; Ministry of Education, Colombo, Ceylon, 1967, 35 pp.

Ashford, T.A. - Contribution of the ACS Examination Committee to Chemical Education; Journal of Chemical Education, Volume 42, 1965, p. 496.

Bloom, B.S. (ed.) - Taxonomy of Educational Objectives: Handbook I, Cognitive Domain; Longmans, Green and Co., New York, 1956, 207 PP.

Bruner, J.S. - The Process of Education; Harvard U.P., Cambridge, Mass., 1960, 97 pp.

Examinations Committee ACS - National Standardization Tests in Chemistry at all levels - High School, College and Graduate Level. 1935 to date.
(May be obtained from Examinations Committee ACS, University of South Florida, Tampa, Florida 33620.)

Furst, E.J. - Construction Evaluation Instruments; Longmans, Green and Co., New York, 1958, 334 pp.

Goodlad, J.I. - The Changing School Curriculum; Fund for the Advancement of Education, New York, 1966.

IUPAC - The Effect of Examinations in Determining Chemistry Curricula (Mathews), Butterworths, 1967.

Krathwohl, D.R., Bloom, B.S., and Masia, B.B. - Taxonomy of Educational Objectives: Handbook II, Affective Domain; David McKay Co., New York, 1964, 196 pp. The second part of the work listed above (see Bloom).

Lindquist, E.F. (ed.) - Educational Measurement; American Council on Education, Washington D.C., 1966 (sixth printing), 819 pp.

Ministry of Education (Department of Examinations), Ceylon - Amended Syllabuses in Science Subjects, Ceylon General Certificate of Education (Ordinary Level) Examination; 1965 (includes sample question papers).

Schools Council Publications, H.M.S.O., London -
1. Examinations Bulletins, particularly Nos. 1, 3, 4 and 8.
2. Working Papers (all available).

The Schools Council - Educational Implications of Social and Economic Change, H.M.S.O., 1967.

Taba, H. - Curriculum Development: Theory and Practice; Harcourt, Brace and World, New York, 1962, 526 pp.

Unesco - Meeting of Experts on the Curriculum of General Education: Final Report, Moscow, January 1968; Unesco, Paris, June 1968, 29 pp.

APPENDIX C

LIST OF PARTICIPANTS

CEYLON:

Prof. G.C.N. Jayasuriya, Dept. of Chemistry, Vidyodaya University, Gangodawila.

Prof. M.U.S. Sultanbawa, Dept. of Chemistry, University of Ceylon, Peradeniya.

Mr. B.J.P. Alles, Deputy Director General (Secondary Education), Ministry of Education, Colombo.

Mr. G.N. Edward, Lecturer, Teachers College, Palaly.

Mr. D. Gunaratne, Education Adviser, Ministry of Education, Colombo.

Mr. D.J. Nanayakkara, Education Officer, Ministry of Education, Colombo.

Mr. P.A.P. Nonis, Lecturer, Teachers College, Maharagama.

Mr. A.M. Ranaweera, Chief Education Officer (Science), Sc/Maths Project, Ministry of Education, Colombo.

Mr. J. Ratnaike, Chief Education Officer (Curriculum Development), Sc/Maths Project, Ministry of Education, Colombo.

Mr. P. Samaranayake, Sc/Maths Project, Ministry of Education, Colombo.

INDIA:

Prof. M.C. Pant, Head, Department of Science Education, National Institute of Education, NCERT, India.

MALAYSIA:

Mr. Tsang ah Liat, Chief Examiner - Chemistry, Malaysian Examinations Syndicate.

U.K.:

Prof. H.F. Halliwell, Professor of Chemical Education, University of East Anglia (formerly Organizer of Nuffield 'O' Level Chemistry Project). Chairman of Workshop.

	Mr. J.C. Mathews, Examiner, Nuffield Chemistry (author of IUPAC Report on Examinations).
	Prof. J.F. Kerr, Professor of Education, University of Leicester.
	Mr. D.G. Chisman, Secretary, IUPAC Committee on the Teaching of Chemistry; Assistant Director, CREDO. Secretary of Workshop.
U.S.A.:	Prof. T. Ashford, University of South Florida (Chairman, ACS Committee on Examinations in Chemistry).
	Prof. C.M. Wheeler, University of New Hampshire, U.S.A. (Representing N.S.F. India.)
U.S.S.R.:	Prof. S.A. Balezin, Unesco Adviser, NCERT, New Delhi, India.
Unesco:	Dr. R.H. Maybury, Division of Science Teaching, Unesco, Paris (Project Officer, Unesco Chemistry Project in Asia.

II SUPPORTING EVIDENCE

1. The Educational System of Ceylon from the point of view of science education

by H.F. Halliwell

What follows is generalised, sure to be inaccurate in a number of specific cases, but gives an adequately accurate working picture. The Ceylonese educational system is closer in its structure to that of the U.K. than to that of many other European countries or to that of the U.S.A.

The main points of importance can be communicated by two figures and accompanying notes.

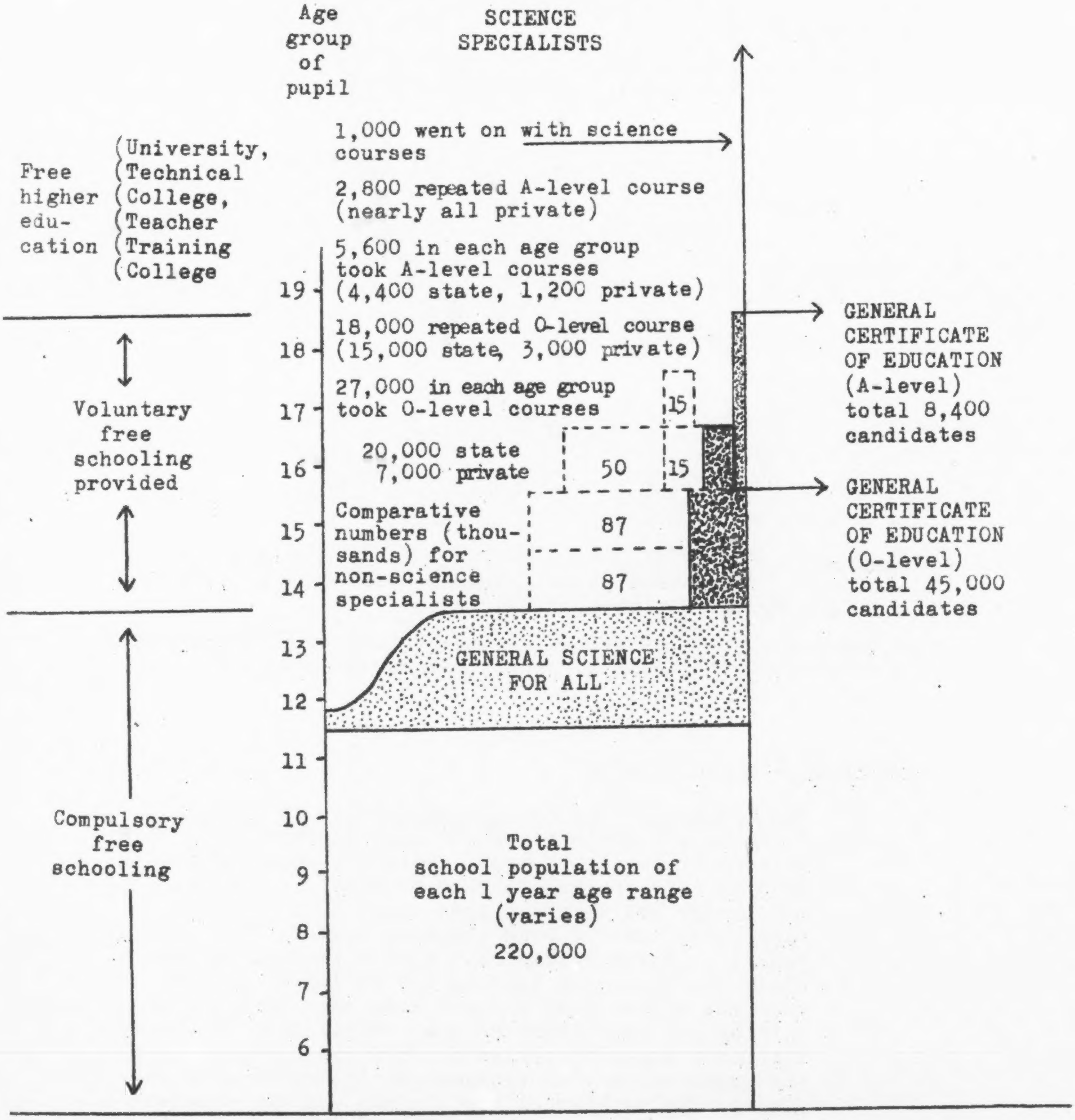

Age group of pupil
SCIENCE SPECIALISTS
1,000 went on with science courses
2,800 repeated A-level course (nearly all private)
5,600 in each age group took A-level courses (4,400 state, 1,200 private)
18,000 repeated O-level course (15,000 state, 3,000 private)
27,000 in each age group took O-level courses
20,000 state
7,000 private
Comparative numbers (thousands) for non-science specialists
15
50
15
87
87
GENERAL SCIENCE FOR ALL
Total school population of each 1 year age range (varies)
220,000
GENERAL CERTIFICATE OF EDUCATION (A-level) total 8,400 candidates
GENERAL CERTIFICATE OF EDUCATION (O-level) total 45,000 candidates
Free higher education
(University,
(Technical College,
(Teacher Training College
Voluntary free schooling provided
Compulsory free schooling
19
18
17
16
15
14
13
12
11
10
9
8
7
6

Figure 1 Structure of school population in Ceylon (1967) showing the population context of the programmes in General Science and in Chemistry, Physics and Biology for O-level and A-level examinations (numbers approximate and rounded from statistics supplied by the Ministry of Education).

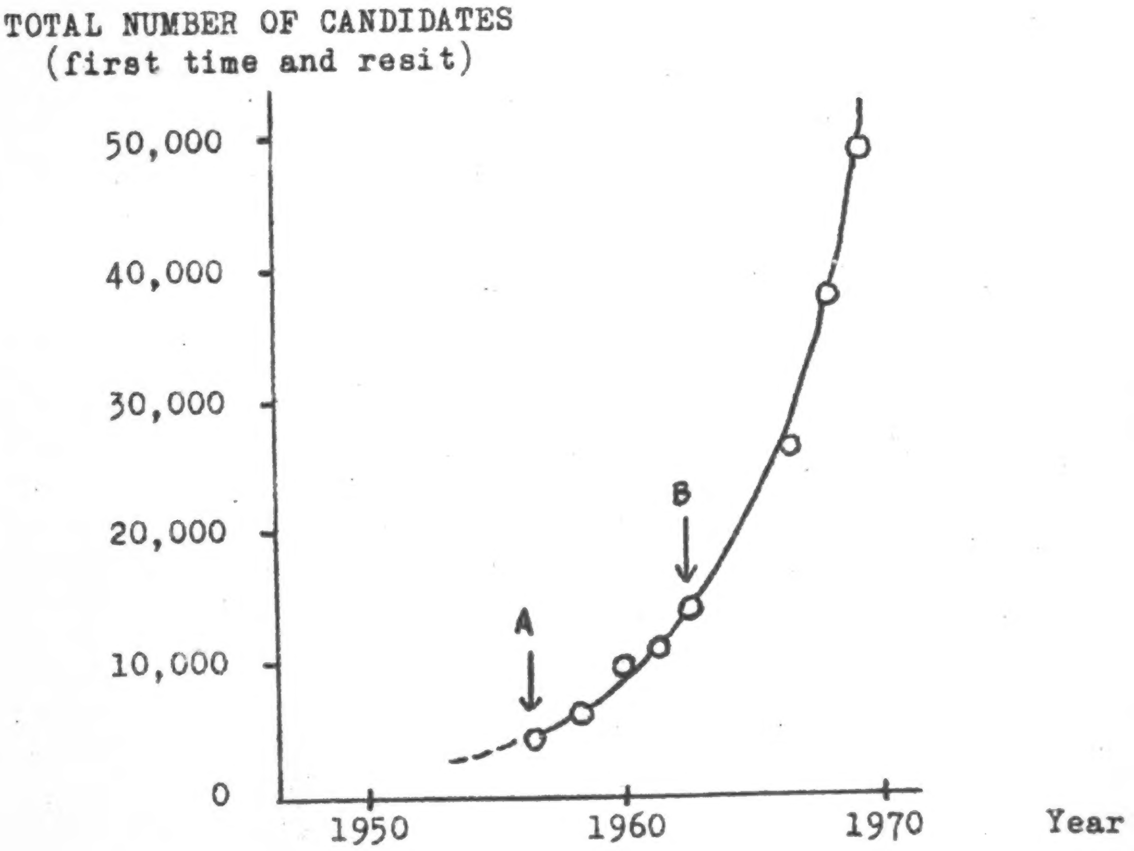


Figure 2 Number of pupils taking O-level Chemistry in successive years (numbers rounded from statistics supplied by the Ministry of Education).

A and B mark the initiation of the General Science programme and the O-level Chemistry programme respectively.

Notes on Figures 1 and 2

(i) The Ministry of Education and Science in Schools. Although all state education, at all levels including university level, is free, there are some private schools (either fee charging or free). Some of the latter are well equipped, have good library facilities, are of long standing and are held in high esteem. Nevertheless, all pupils staying at school after the statutory leaving age of 13½ do so, at whatever school they may be, with the intention of sitting for the Ceylon General Certificate of Education in a range of subjects - and the authority for the establishment of syllabuses, the setting of the papers, the marking of the scripts and the publishing of the results resides with the Ministry. To this must be added the fact that the provision and organising of in-service courses to help teachers meet any changes in syllabus also lies entirely with the Ministry. In Ceylon, the considerable movement for reform in science teaching has also come through the Ministry, namely from the Deputy Director General for Secondary Education and his staff.

In planning, especially in planning work for the senior pupils, the closest co-operation with the country's leading scientists is regarded as top priority. Nevertheless, in schools, in-service training and assistance together with the emergence of new patterns for examinations in the Ceylon General Certificate of Education are the major vehicles for action.

(ii) The Ceylon General Certificate of Education. This examination, like its counterpart in many countries in Asia and Africa, has evolved from, and still strongly resembles, the General Certificate of Education of the U.K. In Ceylon it is held in December. It can be taken at two levels: at the Ordinary Level at the age of 15+ or 16+; and at the Advanced Level at the age of 17+ or 18+. These examinations are nearly always referred to as the O-level and the A-level examinations.

The O-level papers are set and moderated under the direction of the Commissioner of Examinations by a combined group which consists of university members, staff of the Deputy Director General for Secondary Education and members of staff from teacher training institutions. The marking of the scripts is done centrally in the Ministry. Much further increase of the number of scripts shown in Figure 2 will obviously lead to major administrative difficulties. In this connection it must be remembered that the examination in each science subject consists of two papers (one multiple choice of 3/4 hour and one free response of 2 hours). Moreover, answers may be in Sinhalese or Tamil or English, although each candidate must use one language only.

The A-level papers are set by invited university staff and an invited senior university member acts as moderator. The scripts, in this case about 1/10 the number of O-level scripts, are marked on a paid, but voluntary, basis by university staff.

(iii) The pupils

What is the background and intellectual calibre of the pupils taking the examination papers? In seeking an answer one becomes aware that the pupils taking science are influenced by two pressures: selectivity by the educational system and competition from their contemporaries.

Competition can be dealt with briefly but its importance must not be under-rated. It arises from widespread unemployment. The fact is that if a boy obtains his O-levels and then his A-levels in science subjects, then, whether he goes to a university or other institution of higher education or not, there are more opportunities for him to get a job. If he obtains O-level and even A-level passes in arts subjects he may well find himself unemployed unless he has strong family influence backing him.

To do well in science is, therefore, of importance in a context which is much wider than an educational one.

Selectivity by the educational authority also becomes increasingly severe as the pupil advances. To get into an O-level class the pupil must have done well in his mathematics, in general science and in his first language (Sinhalese or Tamil, or occasionally English). Parental wish can also play a part at this stage, and as can be seen from Figure 1 about 10% of the country's age group joined O-level science classes in 1967. Because the selection is partly on academic ability and partly on parental wishes, there is a fairly wide range of ability at the O-level examination.

To go into an A-level science class, however, a pupil must have passed the O-level examination in his first language, in mathematics, and in four other subjects. Further, among these six passes there must be three which he has passed at credit level - two of which must be in the science subjects he proposes to take at A-level. This high standard for entry into the A-level class is one of the reasons why such a high proportion of those taking O-level re-take the examination a year later. The other reason is that if they have failed first time they try to pass on the second occasion for the economic-sociological reasons given above.

To get into a university a pupil must achieve at least 3 credits (not just passes) in his four A-level subjects and also have reached an arbitrarily set level in the aggregate of marks for all four subjects.

These requirements must be set against a background of provision for learning. In O-level classes a pupil has 5 x 40 minute periods per week in each science for two years, and at the end he must take papers in all three sciences (physics, chemistry, biology) and in mathematics. In A-level classes, the time allowance averages 8 x 40 minute periods per week in each of his four A-level subjects - one third to one half of this time being spent in a laboratory. The usual combination of subjects for A-level are: physics, chemistry, pure mathematics and applied mathematics, or chemistry, botany, zoology and physics.

In 1966, 53% of those sitting O-level chemistry passed, and 22% obtained a credit. About 15% did well enough in their chemistry and other subjects to enter A-level classes.

(iv) The teachers

Teachers of A-level classes are all graduates in science subjects. Some will have had professional education training.

Teachers of O-level classes are, by a large majority, non-graduates, but have been professionally trained at one of the two Ceylonese teacher training colleges.

The Ministry, however, has an extensive in-service training and aiding programme. Although each year there is usually only one vacation course in science lasting 2-4 weeks, there is a well established system of one-day study circles taking place at weekends under the direction of the staff of the Deputy Director General for Secondary Education. These study circles are held all over the island, number about 40-50 in a year, and have been running now for four years. This means there is a very close link between teachers on their home ground and those in the Ministry responsible for planning and implementing new schemes.

(v) Some special problems

One of these is common to many countries but perhaps particularly to Asian countries. Centuries of tradition have held that the function of the pupil was to learn at the feet of the teacher whose business it was to know. Any new relationship between pupil and teacher which in any implies that the teacher cannot supply the answer for the child to learn is regarded with suspicion amounting in some cases almost to contempt by many parents and pupils, and even teachers. There is thus an inherent resistance to new modes of learning based on co-operative exploration and it is certainly necessary for any new proposals to be accompanied by extensive guidance for teachers.

But perhaps one of the great problems in Ceylon's education system is that of language, a problem also met in some European and other Asian countries. The language of pupils in Ceylon is approximately:

75%	Sinhalese
20%	Tamil
5%	English

But the special and crucial point is that on the whole neither group can speak the other native language. Very few, even educated, Ceylonese can converse fluently with a countryman in the other native tongue; communication between Tamils and Sinhalese is predominantly through English. It is not without significance that the first version of the new chemistry programme is printed in English although Sinhalese and Tamil versions are being prepared; the problem arises, however, of new technical words in the native language. Examination papers have to be printed in three languages, but those marking one set of scripts on the whole could not mark the other. The situation with respect to textbooks is not an easy one, especially at A-level, for the cost of producing a book that can only be read by a small group within the island and by no one outside it is very high. Nevertheless progress has already been made, as will be discussed in the second section of this Working Paper.

Lastly, any country will be confronted with the cost of equipment, and Ceylon's educational budget in absolute terms is not big. In general science, where the number of classes and of pupils is high, improvisation is essential. At O-level, improvisation is still necessary, but special apparatus and chemicals are available as required. The policy at A-level is that what is necessary for sound learning must be supplied - but reform at A-level is only just being planned.

On the other hand it must be realised that scientific equipment and chemicals, at least at the beginning of the task of coping with the development, need to be purchased from abroad (Europe is the chief source) and transported by sea. This means the use of foreign currency and delays in delivery. Unless the item required is held in stock there may be up to a year's delay in a school's obtaining it, a situation aggravated by the closing of the Suez Canal. The Ministry has begun a scheme for producing locally must of the scientific equipment needed by schools. About 18 miles out of Colombo, building has begun on a site for an Institute of Science Education. Part of the site is being developed as a residential centre for training science teachers and for housing curriculum reform groups. Part of the site comprises drawing office, production workshops equipped with modern machinery and a distribution centre for apparatus designed and tested in the other part. This production centre is already in initial stages of operation. Such an imaginative step, possibly vital to Ceylon's development of science education - especially if there is to be a big increase in the number of pupils in O-level and A-level classes - is almost certain to clash with business interests of foreign suppliers.

Some aspects of the development of the new schemes for teaching chemistry in Ceylon

What follows is not an outline of the history of the Curriculum Reform Project, but a series of comments on some of its problems and some of its successes and failures, attention to which is thought likely to be useful to any group involved in a similar effort. A project focussing on O-level work will have a different set of problems from one concerned with A-level. The latter deals with a group of pupils who are more mature and of a narrow range of high ability. Moreover, the content is much more sophisticated and, for many pupils, is transitional to university work. The problems of A-level work are chiefly concerned with establishing, both in the field of planning and in the area of in-service teacher training, close and fruitful co-operation between university world and school world. Ceylon is only just embarking on its A-level project. This part of the paper is therefore limited to the O-level project, begun in 1961 and now in its third year of trial in schools on a scale large enough to be examined nationally.

It is convenient to consider the problems in the framework of the four steps which will need to be taken:

1. Formulating the educational AIMS of the Project.
2. Planning and carrying out lines of ACTION thought to be necessary to achieve the aims.
3. Setting up machinery for valid and reliable ASSESSMENTS of the action in the light of the aims.
4. Establishing permanent machinery for RE-APPRAISAL on the basis of feed-back and the results of assessment.

AIMS: These are formulated at different levels. There are the general, that is politico-sociological, statements of intentions made by Ministers of Education and the highest directorate level of administrative services; there are the directions given to teachers in terms of programmes for schools; there are the intentions (often imprecisely formulated) of teachers in the classroom; there are the definite demands made on pupils in examinations. That this translation of objectives from general terms through to pupil behaviour should be checked for internal consistency might well be thought to be an obvious necessity. It is not commonly done explicitly. How far it has been done in Ceylon may be judged from what follows.

The general aims that were behind the initiation of the General Science Project in 1957 and of the O-level Specific Sciences begun in 1961 were also stated in the report of the National Commission on Technical Education, July 1963, Sessional Paper X of 1963.

Para. 76. In a developing country, persons having the necessary aptitudes and ability should be selected and trained to make their fullest contribution to the various fields of industry, agriculture and commerce. The purpose of education is not merely to mould the personality of the student in order to fit him into the society in which he lives, but also to develop his innate talents and skills, so that he will be an asset to the nation. Technical education merits special attention in view of its vital relationship to national development.

Para. 78. The education provided at present is mainly academic, and is influenced largely by the requirements for entry to a university. The World Bank Mission has drawn attention to the fact that in the family the children, especially among the more prosperous classes, are excessively sheltered. Initiative, self-confidence, experimentation, inquiring habits of mind, are not encouraged from early childhood. The tendency in education has been to stress the encyclopaedic and purely academic aspect of education, rather than skills and independent solution of new problems. It has served to create a prejudice against manual work and the acquisition of much needed skills; and has thus obstructed economic development. Book learning and memory rather than reasoning ability have been stressed, even in village schools, where children come from predominantly agricultural communities. (Page 23)

At a subject stage of translation, now directed much more precisely at teachers, we find:

> "The pregnant social changes that had taken place in Ceylon, some of which have been indicated above, made it clear that their effects - the changes in the nature, composition, aspiration and function of the G.C.E. 'O' Level population - demanded a change in the educational objectives hitherto assumed - objectives which involved not only the knowledge and skill components of education, but also the component of attitudes."

(Ceylon Science Teaching Development Programme; a report prepared for Unesco, Department of Education, Ceylon, 1963.)

These directives now in the form of questions are repeated in slightly altered form in the detailed printed version of the O-level Chemistry Schemes (Syllabuses of Introduction and Schemes of Work in Science Subjects for G.C.E. Classes, Department of Education, Ceylon).

Nevertheless it is pertinent to ask for operational definitions of some of the phrases used. For example, what is it that a "scientifically literate citizen" can do that distinguishes him from a non-scientifically literate citizen?

Until this is done it is doubtful whether we have any basis for judging whether the examinations which pupils have to take can find out whether they are scientifically literate or not.

Lastly, the Specifications for Examinations have been formulated by the Ceylonese Ministry. Here the influence of "The Taxonomy of Educational Objectives", Bloom et al. (New York, 1956) is acknowledged. The Specifications are stated in the report prepared for Unesco (supra).

The details of the way these specifications are used to plan an examination paper, and the statistical methods by which the validity and reliability of the assessment is checked, are given in Part III - Background Papers, 2 and 3. This aspect of the work has been done with great care and it is possible therefore for the Ceylonese Ministry and for outside observers to judge to what extent the objectives are being achieved.

OBJECTIVES 1, 2, 3, 3', 4

An Explanatory Note

In this analysis objectives are classified using a modified form of "The Taxonomy of Educational Objectives", designed by Bloom et al. for the Cognitive Domain (Bloom, Benjamin S., et al., 1956: A Taxonomy of Educational Objectives; Longmans Green and Co., New York).

1. This category includes test items which require a response involving primarily the RECALL of specific facts, terminology, conventions, trends, classifications criteria, methodology, principles, generalisations, theories, etc.

2. Test items in this category demand a response calling for more than simple recall. An element of COMPREHENSION is required and may be of the following types:-

   translation of mathematical verbal material into symbolic statements and vice versa,

   interpretation of data, statements, etc. (including simple explanations, summaries, etc.),

   extrapolation from data, statements, etc. (i.e. extension of trends beyond the given data, determination of effects, corollaries, etc.

3. Test items in this category will demand responses involving the APPLICATION of abstractions in particular and concrete situations. The abstractions may be in the form of general ideas, rules of procedure, general methods, technical principles, theories, etc.

3'. These test items will require responses such as in 3, but will include numerical CALCULATION.

4. Test items in this category will include items usually classified in Bloom 4, 5 and 6, and will involve responses calling for ANALYSIS, SYNTHESIS and EVALUATION.

   Among others this will include the following:-

   recognition of explicit and implicit relationships, structure, etc., recognition of unstated assumptions, discriminations of facts from hypotheses, organisation of ideas, statements, data, etc., proposition of methods of testing hypotheses, theories, etc., design of simple experiments, recognition of logical accuracy, consistency, etc., comparison of theories, generalisations, etc.

Subject: Chemistry

TABLE OF

CONTENTS	QUESTIONS (Type I)																								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1) Nature and scope of chemistry																									
2) Substances and their properties																									
3) Solutions												X					X								
4) Mixtures and compounds	X																								
5) Physical and chemical changes		X																							
6) Burning																		X							
7) Air	X																								
8) Oxygen		X						X																	
9) Combustion, conditions for			X															X							
10) Oxidation and reduction I				X															X						
11) Action of acids on metals															X										
12) Hydrogen																			X	X					
13) Action of water on metals																			X	X					
14) Oxidation and reduction II					X	X		X																	
15) Atomic theory							X								X				X	X					
16) Classification of reactions								X																	
17) Chemical calculations							X		X													X			
18) Atomic structure																							X	X	
19) 20) Carbon chemistry										X															X
21) Ionization											X					X									
22) Corrosion																									
23) Chemicals from the sea												X													
24) Chemicals from the earth													X												
25) Chemicals from the air																									
26) Energy changes																									
27) Radiochemistry														X											
1	X	X	X	X	X	X	X	X	X	X	X	X	X	X											
2															X	X									
OBJECTIVES 3																	X	X	X	X	X	X	X	X	X
3'																									
4																									

SPECIFICATIONS

															QUESTIONS (Type II)										OBJECTIVES				
26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	1	2	3	4	5	6	7	8	9	10	1	2	3	3'	4
																									X		X		
							X																		X				X
							X																		X				X
																											X		
																									X				
																		O							XX			O	
															O						O				X		XO		O
																O									X		XO		
								X																		X			X
																	O										XX	O	
																											XX		
																									XXX		X		
									X	X													O		XO	X	XX		XX
											X														X				X
					X	X												O							XX		X	OXX	
												X							O								XX		XO
																				O	O				X	O	XO		
X														X							O				X	X	XO		X
	X																					O					XO		
																							O	O	XO	O			
																								O	X	O			
		X																						O		O	X		
			X	X									X								O						XX		X
																									X				
																							O						
																				O				O					
X	X	X	X	X												O	O				O	O							
					X	X												O											
							X	X	X	X	X	X	X	X	O				O										

Examination: G.C.E. 'O' Level

Amended syllabuses - Photo-type paper. SA 636(Sc)

ACTION: The steps necessary to bring a scheme into being in a classroom have roughly followed the following lines:

1. planning and decision making, and establishing a fairly permanent reform group;
2. writing, collecting feed-back, and re-correcting of detailed schemes of work;
3. the setting up of extensive and continued in-service and preservice teacher training;
4. the provision of adequate learning situations and resources;
5. the establishment of a valid and reliable instrument for assessing pupils' achievement.

What has the Ceylonese group done about this?

The Science Teaching Project has been centred in the Ministry under the direct and extensive guidance of the Deputy Director General for Secondary Education. He and the staff concerned have had close links with similar work in other countries either by going to these countries to work intimately with the proposals or by the bringing of overseas experts to Ceylon. The extent of the careful look at other people's efforts can be judged by the following facts. Two of the staff spent one year in the U.K. (another three spending six weeks there); sixteen spent one year in the U.S.A. and two spent short periods in India and Australia; two, closely associated with the Chemistry Project, spent a year in Bangkok with the Unesco Chemistry Teaching Pilot Project for South-East Asia, and one of these has continued this close association with this Asian project for a second year.

In the reverse way, groups of leading science teachers from the U.K. were invited to take part in inspectorate discussions and teacher-training courses, and leading figures associated with reform in the U.S.A. and Australia have been invited to Ceylon.

In this way the Ceylonese group were well informed of modern trends in other countries and had time to translate these into details which were appropriate to their own needs. The basis for the preparation of detailed schemes of work was the writing of a preliminary draft, trial and feed-back, rewriting, further trial and feed-back and finally extensive printing and distribution followed by an intensive system of trial and feed-back lasting over 1½ years.

Such work is now being continued by the science supervisors in consultation with the writing group. Because the writing group had so much contact with movements in other countries, were themselves experienced teachers and had already formulated a set of objectives which acted as criteria for selection or rejection, they were able to write a draft scheme that had programmed learning as a basis and was essentially viable. The feed-back was limited to begin with. Groups of six or seven teachers were asked to try out small sections, and more extensive areas were tried out with teachers at vacation courses. The latter were very valuable in that not only was feed-back obtained, but the teachers themselves were introduced to and participated in the new work that was yet to come.

After the writing of the second draft there were more extensive trials lasting over one year. Now it was arranged that a range of types of schools and pupils was covered, the writing group themselves going into schools and teaching as full time teachers.

During the vacations, discussions lasting one or two weeks were held with selected senior science teachers.

Finally a printed version was published along with "Amended Syllabuses in Science Subjects" (Department of Examinations SA 636(Sc), 1964). Following this, very extensive in-service trials were held by the working group every weekend for 1½ years. In this way in the course of a year almost all teachers in the island were contacted. The dedication necessary and the amount of labour, unrelieved by any break, which was involved is obvious to say the least.

As a result of such interchange of experience, certain areas of the printed scheme are now under revision. The setting up of in-service training has been covered in commenting on the writing and collection of feed-back - in fact these two aspects are not separable. Pre-service training which supplies the majority of (non-graduate) teachers for O-level work is under a separate branch of the Ministry. The work of teacher training and of curriculum reform should be closely co-operative. In Ceylon the two are not yet as closely knit as they are in some other countries.

In turning to the setting up of adequate learning situations for science one is chiefly concerned with finance: supplying what is needed by a school (and conversely seeing that what is supplied is in fact needed) and building laboratories.

Ten years ago there were about 166 school laboratories in the island, chiefly in the old-established and private schools. Now there are about 700. Ten years ago in most schools science teaching was chiefly a matter of use of blackboard and a few demonstrations. Now the inspectorate report that in most schools pupils are doing experiments at least once a week. In the last ten years, too, there has been set up a centralised store of equipment, a system of requesting supplies, and a centralised repair and workshop unit. The cost of this change, quite apart from the cost of buildings, is seen from the following figures for state schools.

	Equipment	Chemicals
1957	156,000 Rs	52,000 Rs
1963	400,000 Rs	173,426 Rs
1967	1,400,000 Rs	325,605 Rs

A list of minimum equipment has been prepared by the Ministry.

There are about 150 laboratories in which O-level and A-level chemistry is done. These have good facilities and usually one or one and a half laboratory assistants to each laboratory. General science is not taught in these rooms. There are 400-500 laboratories, less expensively equipped, which are used for O-level work only, sometimes for two subjects. Most of these have electricity or gas produced on the premises. Running water is not available in many schools because of the expense of piping it, but much of the work at O-level can be done with a static supply and a system of syphoning.

There is one ARISBA laboratory in use, erected by the U.N. School Building Unit as a prototype for Asia. This has static water, storage cells and portable arrangement for heating, operationally planned arrangement of working places, and storage facilities.

The question of books involves the language problem mentioned before but at the same time provides for the evidence (along with the type of examination questions pupils now have to answer, and together with the increased amount of experimental work) that the new schemes are not just on paper: a differently trained group of young people is emerging.

In the 1940s the principal textbooks in use for O-level in the island were Littler: Elementary Chemistry, and Mee: School Chemistry.

In the 1950s the popular textbook was Wilkins: Fundamental Chemistry (translated into Sinhalese and Tamil in 1956).

In the 1960s Holderness and Lambert's School Chemistry has been predominantly in use.

But the fact is that with the change of programme and type of examination question it is difficult to answer the new type of demands by reading only the books quoted above. Even the content of the new schemes is not covered by those more traditional texts.

The Ministry has therefore had to have new textbooks written and this task has been undertaken by those responsible for the reforms in science education. Unfortunately these new texts are only published in Sinhalese or Tamil. In the opinion of the compiler of this paper, these texts should be translated into English so that those who are considering writing textbooks for their own country can benefit by seeing how ideas from many other parts of the world can be woven into something that is relevant to ones own country.

The characteristics of these new text books are the completely fresh look at content and approach, their inclusion of much of interest from schemes in other countries, the deliberate building into the text on almost every page of little answers and questions to the pupil reader and suggestions as to what he might do at home, and (unusual in a "textbook") chapters on the method of solving problems by investigation, weighing of evidence for hypotheses and putting forward explanations. These chapters take the process of being scientific away from classroom and school into the village, the home, the world of newspapers and personal argument, and then bring it back again to the immediate problem of the chemistry lesson. There is a deliberate attempt here to make the phrase "scientifically literate" take on an operational meaning. It is a bold and imaginative piece of production, yet all the time it has its feet firmly on the ground.

ASSESSMENT: There are two separate areas where a valid and reliable method of assessment is necessary.

The first is in the realms of the Ministry's own activities. It laid down its objectives clearly. Are these objectives being reached? Some of the evidence will be qualitative: the different kind of examination question, the increase in laboratory work in schools, the demand for a new type of textbook. This will give enough indication for further action to be taken, but a detailed statistical assessement of the success of the policy is a highly specialised research project and is not available at the time of writing this Working Paper.

The second area is that of pupils' activities. The instrument for this is usually a public examination. In Ceylon this instrument has been subject to careful scrutiny over seven or eight years. The analysis of papers for content validity and objective validity, the statistical analysis that has been done on the results, are outlined in Part III - Background Papers, 2 and 3.

2. Design for Setting the Sample Question Paper in Chemistry for Higher Secondary School Leaving Certificate Examination

NCERT, India

In order to eliminate the wild role which subjectivity plays in setting, answering and scoring questions, to ensure a better coverage of objectives, a more effective distribution of questions over the whole syllabus, inclusion of a variety of questions and thereby to do away with selective study and cramming on the part of pupils, the following design for setting the question paper in chemistry is suggested.

The design is only a suggestion, and leaves scope for necessary alterations according to the needs and nature of the subject matter. The suggestions are intended to be followed in spirit and the weightages proposed to different items may be considered as only approximate. The setter can at his discretion, therefore, feel free to make changes in respect of the weightages, if necessary.

I Weightage to objectives

In order that the questions framed may not be based on just one or a few objectives, it is also imperative to determine the weightage to each objective and thereby distribute questions over all the objectives in proportion to their importance.

In view of the foregoing considerations, a scheme for the weightage to different objectives for this paper is suggested below.

TABLE I

Weightage to Objectives

S. No.	Instructional objectives	Marks	Percentage
1	Knowledge	48	48
2	Understanding	37	37
3	Application	15	15
	TOTAL	100	100

II Weightage to content areas

Towards obtaining a wider and effective coverage of the syllabus it is suggested that an earnest attempt be made to distribute the marks alloted over different parts of the syllabus.

TABLE II

Distribution of Marks over Major Areas of Content

S. No.	Areas of Content	Total Marks	
		Major areas of content	Syllabus Divisions in Question Paper
1	Laboratory processes	2	2
2	Chemical and physical changes	8	8
3	Atomic weight, molecular weight, laws, etc.	6	6
4	Solutions	3	3
5	Halogens	6	6
6	Acids, bases and salts	8	8
7	Nitrogen and phosphorus	10	10
8	Sulphur	10	10
9	Na, Ca, Mg and Si	7	7
10	Metals	20	20
11	Carbon and hydrocarbons	15	15
12	Qualitative analysis	5	5
	TOTAL	100	100

III Weightage to forms of questions

It is common experience that the inclusion of only a few essay questions in the question papers encourage guess-work and cramming of set answers to selected questions. It is therefore suggested that the question paper may include all the four types of questions, viz. the essay type, the short answer type, the very short answer type and the objective type.

A suggested weightage to different forms of questions is given below.

TABLE III

Weightage to Different Forms of Questions

S. No.	Forms of questions	No. of questions	Marks	Percentage
1	Essay	5	25	25
2	Short answer	25	50	50
3	Objective	25	25	25
	TOTAL	55	100	100

IV Difficulty level, language and arrangement of questions

One of the important functions of examinations is to discriminate between the low achievers and the higher achievers. A question paper is considered to be a good paper if it discriminates well. For this purpose it is recommended that the question papers may contain in due proportion questions of various difficulty levels ranging from easy to difficult. It may, however, be ensured that they are not too easy nor too difficult for the H.S.S.L.C. level.

The wording of the questions must be sufficiently clear, explicit and unambiguous for each examinee to comprehend the questions in the same way as every other examinee and to be able to know precisely what he is expected to write or draw and of how great a length.

From the psychological point of view and in order to create confidence in the examinee, the questions as far as possible should be arranged according to the increasing level of difficulty. This arrangement may be done within the topic or the content also if necessary.

V Scheme of option

Options as such have no place in a scheme of good examinations, and hence no option is provided in this paper.

VI Scheme of sections and their administration

The question paper may be divided into two sections - Section A may include objective type and very short answer questions; Section B may include essay and short answer type questions.

It is suggested that both the sections be administered simultaneously with clear instructions to the examinees that the answer papers for Section A will be collected promptly at the end of half an hour. After the commencement of the paper, the students will have to be instructed in advance that if they finish Section A before time is called, they may go right on to Section B without awaiting further signal. The time allotted to each section along with other subsidiary details is suggested below.

TABLE IV

Time Allotment to Sections

Sections of the sample question paper	Form of question	Total marks	No. of questions to be attempted	No. of questions to be set	Estimated time for answering in minutes
Section A	Objective	25	25	25	30
Section B	Essay and short answer	75	30	30	150

VII Directions

Adequate directions regarding the number of questions to be attempted, allotment of time to different sections and other necessary instructions may be given clearly in the question paper.

VIII Model answers, scoring key and marking scheme

It is suggested that model answers and marking schemes should be prepared along with the question papers. This will increase uniformity in scoring and will reduce to a great extent the chance of inclusion of vague questions requiring answers of an undefined scope. This will also check possible errors (numerical and others) that sometimes unwittingly creep in. This suggestion should, therefore, be invariably adhered to.

The design suggested here is only a first step towards improving examinations. It may require modifications after it is given a trial for two or three years.

V

SAMPLE QUESTION PAPER

i. Blueprint

S. No.	OBJECTIVE CONTENT	KNOWLEDGE			UNDER-STANDING			APPLICATION			SKILL			TOTAL
		E	S	O	E	S	O	E	S	O	E	S	O	
1	Laboratory processes	..	..	..	..		1(1)	..	..	1(1)				2(2)
2	Chemical and physical properties	..	2(1)	1(1)	..	2(1)	2(2)	..	..	1(1)				8(6)
3	Atomic weight, molecular weight, laws, etc.	..	4(2)	..	..	..	..	..	2(1)	..				6(3)
4	Solutions	..	..	..	..	2(1)	1(1)	..	..	..				3(2)
5	Halogens	5(1)	..	..	..	..	1(1)	..	..	..				6(2)
6	Acids, bases, salts	..	2(1)	1(1)	..	2(1)	1(1)	..	2(1)	..				8(5)
7	Nitrogen and phosphorus	5(1)	..	1(1)	..	2(1)	1(1)	..	..	1(1)				10(5)
8	Sulphur	5(1)	2(1)	..	..	2(1)	..	..	..	1(1)				10(4)
9	Na, Ca, Mg and Si	..	2(1)	..	..	2(1)	1(1)	..	2(1)	..				7(4)
10	Metals	5(1)	4(2)	2(2)	..	4(2)	2(2)	..	2(1)	1(1)				20(11)
11	Carbon and hydrocarbon	5(1)	2(1)	..	..	6(3)	2(2)	..	..	..				15(7)
12	Qualitative analysis	..	..	..	..	2(1)	1(1)	..	..	2(2)				5(4)
	Sub-total	25(5)	18(9)	5(5)	..	24(12)	13(13)	..	8(4)	7(7)				
	Total	48(19)			37(25)			15(11)						100(55)

Note: Figures within brackets indicate the number of questions and figures outside brackets indicate marks.

Summary:			
	Essay (E)	No. 5	Marks 25
	Short answer (S)	No. 25	Marks 50
	Objective type (O)	No. 25	Marks 25

Scheme of Sections: Section A - Objective questions
Section B - Short answer and essay question

3. Notes on the Assessment of Practical Work by Teachers: 'A' Level Chemistry, Nuffield Scheme, U.K.

by J.C. Mathews

The assessment of practical work by teachers has been undergoing trials since September 1966. The purpose of these notes is to explain this form of assessment to teachers who are about to start the Nuffield 'A' level chemistry course in 1968.

The advantages of assessment by teachers

1. The assessment can be done by those who actually observe the students at work over a long period of time, so that practical skills, abilities and aptitudes, which are not displayed in written work can be taken into account. In this way the objectives of the assessment can be made wider than those of a practical examination assessed on a single occasion on written evidence alone.

2. Assessment on several occasions is necessary to cover adequately the variety of experiments and techniques which comprise the practical work of the course; and the effect of chance failure or success is less.

3. The assessed practical work need not be limited to those experiments which can be readily administered to a large number of candidates in a limited time. The restriction of practical examinations to such experiments could have an undesirable effect on the variety of practical work in the schools taking the Nuffield course. The form of assessment set out in these notes should lead to a greater freedom of choice of practical work in schools.

4. This scheme does give teachers the opportunity to become involved in the assessment of their students. It could lead to a greater awareness of the objectives of practical work and a situation in which teachers and Examiners work together towards these objectives.

The outcome of the school trials

During the two years of trials, eight experiments, all of them part of the basic course, were chosen by the Project's Headquarters team for assessment. Guidance was sent to teachers on the conduct of the experiments, the objectives of each, and the manner in which marks could be awarded. The results for each student were entered on a record card and the record cards were reviewed by the Examiners at the end of the first and second years. The main conclusions were as follows.

1. Teachers tended to use a restricted range of marks, and a candidate's total mark was rarely less than the half mark. The mean mark was correspondingly high. From an analysis of the marks, coupled with the direct observation of practical classes by the Headquarters Team, it can be inferred that the spread of ability is less in practical work than it is in written work, and it is unlikely that a student can be said to have failed in his two years of practical work.

2. The objectives of practical work were the subject of much debate. Finally, all the teachers who took part in the trials were invited to state what they considered the objectives to be. The objectives and their weighting, which are set out later in these notes, follow closely the outcome of this inquiry.

3. Many teachers felt that the detailed instructions, given by the Headquarters Team for each of the assessed experiments, unduly restricted them in their assessment. It became clear that no single set of instructions would meet the wishes of all the teachers, and it was decided that the proper course was to give as much freedom as possible to the teachers in the choice of experiments for assessment and in the conduct of the assessment. Some uniformity is necessary, however, in the objectives, the areas of work, the type of experiment, and the form in which the results are recorded.

4. It has been a constant policy of the trials that the assessment should be conducted on normal school work and be as unobtrusive as possible, but often there was a noticeable difference in the attitude of a class to the work when they knew it was to be assessed. Some teachers welcomed the more diligent work, others regretted the more formal atmosphere.

5. Teachers were rightly concerned lest the raw mark which they awarded their students would be the one used in the final 'A' level mark. The raw marks were used by the Examiners to determine only the distribution of marks within a teaching set, and this policy will continue.

6. Possibly the most difficult decision which had to be made was whether there was sufficient difference in the general standard of practical work between teaching sets and between schools to call for the use of a moderating instrument in order to make adjustments to the marks. For the time being, it seems that some form of moderation is desirable, but the matter will be reviewed from year to year.

7. Teachers found that those objectives which could be marked from written evidence were easier to assess than those which called for direct observation by the teacher during an experiment. It was found to be very difficult to assess more than

one of the latter during any one experiment. It is hoped that assessment by direct observation will become easier as teachers become more familiar with it and can choose those experiments in which they know they can pick out particular faults and virtues without interfering with the normal teaching work.

The Form of Assessment after 1968

The objectives

The following objectives, with weighting given as a percentage, are suggested.

1. Skill in observation (25%).
2. Ability to interpret observations (15%).
3. Ability to plan experiments (10%).
4. Manipulative skills (30%).
5. Attitudes to practical work (20%).

In many respects the form of assessment advocated in these notes has still to be evaluated, and it would be wrong to regard this statement of objectives and weighting as fixed for all time. Teachers may wish to modify both the objectives and the weighting, particularly in the finer divisions of the objectives. But for the time being it will help in the evaluation of this form of assessment if teachers can keep as close to the suggested pattern as possible. If teachers feel that they must deviate from this pattern they will still be asked to show on the record cards how they allocated marks to particular objectives.

The objectives 1 - 5 provide the broad basis for the final assessment, but for assessment on particular occasions a finer division is necessary. The following notes on the objectives may be helpful, but teachers should not feel that they are rigidly bound by them.

1. Observation alone has little merit unless it is accurate and relevant. Its assessment leans heavily on a student's ability to report, and it is inevitable that most of it will take the form of awarding marks for written records of observation. But there are two aspects of observation for which teachers can give credit while a conventional practical examination cannot, and it is hoped that neither will be overlooked. The first is oral reporting which is usually more immediate and less rehearsed than written reporting. The second is a student's observation and interest in unusual and unexpected features of an experiment even though they are not relevant to the immediate purpose of the experiment. It will not be easy to put a mark on these two aspects but at least credit could be given in a periodic review and grading of each student's attitude to practical work.

2. The assessment of interpretation is similar to that of observation in that teachers can only judge what students write or say. It will give credit for an analytical approach to observations, including unexpected observations and the ability to comment critically on practical situations.

3. The ability to plan is most easily assessed by asking the students to write their plan and submit it for marking before they start, but this can only be done in those experiments for which information to students on procedure is restricted. It should include the ability to recognize sources of error, and the ability to make predictions about other experiments.

It can be argued that both interpretation and planning can be assessed on written examination papers. This may be so, but a student's interpretation of observations which he has actually made and his plans for an experiment which he knows he must try to put into operation may well be different from his interpretation and planning of unreal situations on paper. The thinking processes may be similar but their context is different, therefore their outcome may be different.

4. Manipulative skills can be assessed by marking the results of experiments, for example yield, purity, and accuracy (accuracy of work rather than accuracy of calculation); they can also be assessed by direct observation of such things as orderliness, methodical working, dexterity and speed. It is desirable that this assessment should not be confined to routine operations; it should include the ability of students to adapt their manipulative skills to new situations and to follow unfamiliar instructions.

It may be possible for teachers to use specific experiments for assessing some parts of manipulative skill, particularly yield, purity and accuracy. However, many teachers feel that a grade based on a general impression of work over a period of say a term or even a year is more appropriate for the less exact aspects of manipulative skill which can only be assessed by direct observation. If periodic grading is used, it is important that teachers should consciously look for evidence of the skills throughout the period and not simply rely on their memory at the end of the period. (A further note on grading by impression is given on p. 74.)

5. For the purpose of this assessment attitudes can be divided into three categories:

(i) Willingness to co-operate in the normal routine of a chemistry laboratory. This will include a knowledge of safety regulations and a willingness to comply with them and the other regulations which are necessary for the efficient running of a laboratory.

(ii) Persistence and resourcefulness - the will to work unaided on set problems and to see them through to a meaningful conclusion.

(iii) Enthusiasm: a commitment to practical work as a worthwhile pursuit without compulsion. This could show itself in suggestions for lines of investigation not specified in the course and in assuming the more active role when working in groups.

A grading at the end of each year, or even at the end of the course, is probably the most appropriate method of assessing attitudes.

Areas of subject matter

The assessment in the future will not be restricted to specified experiments as it was in the trials. Teachers will be free to choose which experiments to use and these experiments need not form part of the published Nuffield course. But in order to get a reliable order of merit it is important that the main areas of subject matter are covered. On each student's record card the examiners will expect to see a record of assessment on at least eight experiments, including at least one from each of the following areas of subject matter.

1. Changes in substances and patterns in changes in substances. (This is a deliberately broad category to include most of the course not covered by 2, 3 and 4 below. Most of the assessed experiments are likely to come in this category.)
2. Equilibria.
3. Kinetics.
4. Energetics.

Types of work

Three main types of work can be identified in the course and the examiners will expect to see each one used in the assessment, but not necessarily with equal weighting.

1. Quantitative work.
2. Preparative work.
3. Qualitative work.

Allotment of marks

The number of occasions on which marks are allotted for practical work is left to the discretion of the teacher provided that the minimum number of experiments is attained. It is desirable that the objectives, with the possible exception of objective 5, are assessed on more than one occasion in order to get a reliable estimate of ability.

Students should be told at the beginning of the course that their practical work will be assessed and that the results will be sent to the examiners. They should also know, in general terms, the qualities for which the teacher will look. They should be told that the assessment will be more obvious on some occasions than on others, but that it will usually take place during normal work and not on special occasions. It may reassure them to know that the average mark for practical work will be higher than that for the written examination.

The examiners will not necessarily expect that all the students in a teaching set will be assessed on the same experiments, but an equal balance of objectives is desirable. Assessment of objectives 1 and 2 will be difficult in experiments in which the students work in groups, but teachers should not feel restricted in the assessment to those experiments which are performed by students working alone: some assessment of objectives 3, 4 and 5 is possible in group work.

During the trials a grading scheme (5-4-3-2-1) was used to allocate marks to a particular objective on a particular occasion, and teachers may find a five point scale useful especially for those parts of the work which are marked by impression rather than by a mark scheme, but they are free to use other schemes if they wish.

There are three main ways in which marks can be allocated and teachers will probably find that they have to use all of them.

1. A mark scheme. This will be most useful when marking written evidence of observation, interpretation, planning and accuracy.

2. Marking by impression on a single occasion. This will be useful for marking evidence which is less precise than that mentioned in 1. For example, a teacher may wish to assess dexterity in handling unfamiliar apparatus - say in chromatography. This can best be done by direct observation of the students at work, and in a large class it is not easy. Teachers should try to assess only one such quality by impression during one experiment. Furthermore, teachers are usually too busy to do more than make a note of those who are doing well or badly; the rest can only be given the mark for average work.

A good deal of thought during the trials was given to the question of whether "average" should be the teacher's estimate of the average for his set or his estimate of the average for 'A' level work in general. The former has the disadvantage that a teacher cannot estimate the general level of practical ability of his set until he has been teaching them for some time. The disadvantage of the latter is that it depends on the teacher's experience of 'A' level work. Since the raw marks will be scaled to a norm anyway, it is their distribution and the internal consistency of the marking which is important. But some uniformity is desirable and it has been decided that teachers should try to judge what is average from 'A' level work in general rather than from the work of their teaching set. Thus:

Excellence of work unusual in an 'A' level student -	5
Above average work -	4
Average work -	3
Below average work -	2
Unusually poor work for an 'A' level student -	1

3. Marking by impression after a period of time. This will apply mainly to attitudes to practical work, but teachers may wish to assess other objectives, particularly some of the less precise aspects of manipulative skill, periodically rather than in single experiments. The period is left to the teacher's discretion; it could be done once a term, once a year or, for some objectives, once at the end of the course. The criteria for awarding a grade are those set out in 2.

Information required by the examiners

Three or four weeks before the written examination the examiners will require the following information from schools.

1. All the record cards of the students arranged according to teaching sets. A sample showing how a record card can be completed is given in Appendix A; this should not be taken to be an example of an ideal assessment.

2. A summary of the marks for practical work for each set, showing the total mark awarded to each candidate as a percentage (Appendix B).

3. Any other relevant information, for example:
   - (i) Outstandingly good or bad performance throughout the course by individual candidates.
   - (ii) Outstandingly good or bad performance by a set as a whole.
   - (iii) Circumstances such as absence or change of school which could affect the performance of a candidate, particularly if it has resulted in the completion of fewer than eight experiments in which assessment has taken place.

The moderation of marks

The present need for moderation of the teachers' marks has already been mentioned (p. 70). For various reasons, including those of ineffectiveness and expense, moderation by visiting examiners, moderation by a practical examination and moderation by performance on the written papers as a whole, have been rejected.

Moderation will take place by the performance of teaching sets as a whole in those questions in the written Papers I and II which will have been designed to test practical work; these questions will form not less than 20% of the two Papers. It seems that performance in these questions is more likely to reflect the general level of practical work of a teaching set than any other moderating instrument which can be conveniently applied.

Three ranges of marks will be allocated, based on the performance of teaching sets as a whole in the written practical questions. To conform with the findings of the trials (p. 70) all three ranges of marks will be narrow, they will overlap, and all three mean marks will be higher than the half mark. For example, in 1968 the maximum mark for the teacher assessment of practical work will be 70, which comprises 14% of the maximum mark for the whole examination and the three mark ranges will be:

Top range (20% of teaching sets), mean mark 57.5, standard deviation 4.
Middle range (60% of teaching sets), mean mark 47.5, standard deviation 4.
Bottom range (20% of teaching sets), mean mark 37.5, standard deviation 4.

This form of moderation will be reviewed from year to year.

In conclusion

This is a first venture in this form of assessment of 'A' level practical work in chemistry by teachers, and in many ways it is still on trial. At first it will no doubt seem strange to both teachers and students, but it is hoped that in time it will become accepted as part of normal Sixth Form work. The examiners will welcome queries, comments and suggestions from the teachers who operate the scheme; they should be sent to Mr. J.C. Mathews, 51 Laverton Road, St. Annes, Lancashire.

Appendix A

A SAMPLE OF A RECORD CARD SHOWING IMAGINARY MARKS FOR A STUDENT

N.B. The reference numbers and titles of experiments are based on the written materials before publication.

Name .... School .... Form and set ......

NUFFIELD FOUNDATION SCIENCE TEACHING PROJECT
CHEMISTRY ADVANCED LEVEL
ASSESSMENT OF PRACTICAL WORK

Duration of course: September 19... to July 19...
Record Card Number ...1...

ASSESSED EXPERIMENTS AND OTHER ASSESSMENTS	OBJECTIVES (showing number) AND MARKS (showing maximum)	
2.2c Properties of chlorides	Observation (1) 6/10	Interpretation (2) 4/5
5.3(b) Iodine/iodate reaction by titration	Accuracy (4) 7/10	
6.1(a) Reactions of some metal elements	Observation (1) 8/10	Interpretation (2) 3/5
7.2(b) Enthalpy of hydration	Comment on sources of error(3) 3/5	
9.3(b) Preparation of phenyl benzoate	Yield (4) 2/5	Purity (4) 4/5
-- General impression oof manipulative skill (year end)	Orderliness and methodical approach (4) 4/5	Speed (4) 3/5
-- General impression of attitude to practical work (year end)	Routine (5) 4/5	Persistence (5) 4/5

SUMMARY

....
Signature of teacher

	OBJECTIVES					TOTAL
	1	2	3	4	5	
Marks						
Scaled marks						

Name .... School .... Form..........

NUFFIELD FOUNDATION SCIENCE TEACHING PROJECT
CHEMISTRY ADVANCED LEVEL

ASSESSMENT OF PRACTICAL WORK

Duration of course: September 19... to July 19...
Record Card Number ...2...

ASSESSED EXPERIMENTS AND OTHER ASSESSMENTS		OBJECTIVES (showing number) AND MARKS (showing maximum)		
11.2(b)	Determination of strength of hydrogen bond	Planning (3) 5/10	Interpretation (2) 3/5	
13.3	Solubility product of silver bromide	Accuracy (4) 4/5		
13.6(d)	Prediction of redox reactions	Prediction (3) 3/5	Observation (1) 6/10	
15.2	Rate of reaction tin and iodine	Recognition of errors (3) 4/5		
16.1	Reaction of thionyl-chloride with hydrates	Observations (1) 4/5	Interpretation (2) 3/5	
17.3	Chromatographic separation of amino acids	Dexterity (4) 3/5	Observations (1) 4/5	Interpretation (2) 2/5
--	General impression of manipulative skill (year end)	Orderliness (4) 4/5	Speed (4) 4/5	
--	General impression of attitude to practical work (year end)	Routine (5) 4/5	Enthusiasm (5) 4/5	

SUMMARY

	OBJECTIVES					TOTAL
	1	2	3	4	5	
Marks	28/40	12/25	15/25	32/45	16/20	109/160
Scaled marks	18/25	7/15	6/10	21/30	16/20	68/100

....
Signature
of Teacher

Appendix B

A SAMPLE OF THE FORM OF THE SUMMARY OF ASSESSMENT FOR A TEACHING SET

School .... Form and Set ....

NUFFIELD FOUNDATION SCIENCE TEACHING PROJECT
CHEMISTRY ADVANCED LEVEL

SUMMARY OF ASSESSMENT OF PRACTICAL WORK

Duration of Course: September 19... to July 19...

Candidate Number	Name and initials	Total Mark /100	Leave this space blank

.... Date ....
Teacher's signature

III BACKGROUND PAPERS

1. Evaluation in Chemistry within a Second Level General Education Programme

by J.P. Alles

1.0 The main purpose of the paper and some preliminary comments

This working paper is prepared for a conference relating to evaluation in chemistry within second level general education programmes. It is presented as a basis for discussion and study.

The main purpose of the working paper is the presentation for discussion of the following aspects of the theme of the conference.

1.1 The learning and teaching of chemistry are one component of a second level general education programme. The total second level general education programme will include, in addition to other science subjects, such as physics and biology, several other major components, such as first language, second language, mathematics and so on. Hence, "Evaluation in Chemistry" is one aspect of the larger problem of "Evaluation within Second Level General Education Programmes". Consideration and critical study of evaluation in chemistry, to be complete, may demand a consideration of the general problems of evaluation within second level general education programmes. It may be appropriate to discuss whether all the issues in evaluation in chemistry are specific to chemistry, or whether many are only special cases in the fields of chemistry of a general class of problems of evaluation at second level in general education.

1.2 In the design and implementation of educational programmes many major processes are involved, such as the following:

   i. Analysis of the situation within the educational system as a whole, isolation of significant issues and professional policy decision making.

   ii. Planning and programming of educational activities.

iii. Communication of the planned programmes to executional levels along with relevant resources allocations.

iv. Supervision, assistance and guidance in implementation.

v. Evaluation and assessment in the context of the specified goals.

It will be seen that evaluation and assessment are but one of the major processes taking place within an educational system. Hence, evaluation in chemistry needs to be viewed from this "process standpoint" as well, and should be considered in the perspective of the system as a whole.

1.3 Evaluation in chemistry viewed from these two standpoints will be seen to have many unresolved or partially resolved problems. It may be very important not only to isolate the fruitful and valuable techniques and approaches now available, but also to isolate and to recognise the major unresolved problems which may require exploratory work in the immediate future.

It is recognised that the participants at the conference are primarily concerned with the subject at the level of design, and the emphasis throughout is essentially on recognising useful concepts, techniques and criteria of judgment and decision that are available, and the recognition of new areas for action research.

A working assumption in relation to the entire paper is that evaluation, as a process, is perhaps still in need of major development; and evaluation in chemistry in particular may see substantial progress in the years immediately ahead.

It is relevant to remark that, to those who are not directly concerned with design and innovation and path-finding activities in the fields of evaluation, points 1.1, 1.2 and 1.3 above may have but little relevance. Some may even argue that not only are the viewpoints suggested of limited relevance, but also that the attempt to see evaluation in chemistry as suggested may be positively "confusing". This may be true. The paper will have relevance only in the context of a "design and innovation-oriented" viewpoint.

In order to recognise the status of evaluation in relation to classroom or school activities, questions such as the following need to be considered:

What type of activity and teaching programme goes on within the school?

Where does one get statements and descriptions (specifications) of the school's programme of work?

What are the aims and purposes that the school seeks to achieve? In regard to these what significant problems exist?

What staff are charged with recognising and formulating these problems? Is there an active programme of work directed towards the solution of these problems?

Who decides what goes on in a school and who plans out in general terms, and later in detail, the programmes of work and study within the school? Who assist and guide principals, teachers and pupils at work?

Are there procedures and guidelines laid down for finding out whether the school is achieving the purposes that it has set itself to seek?

How is the school associated with its immediate community? How is the school associated with the larger community, the nation?

What methods and techniques and procedures are available for finding out whether the pupils are in fact developing adequately and learning effectively?

What administration and management techniques have been established so that the work of the school is linked to the entire hierarchy of activities directed towards the development of the community and the nation?

All these questions and many others of a similar nature are relevant in the context of a discussion on evaluation. Depending on the nature of the task in hand, some of them become more important than others. In this paper it is suggested that staff engaged in "Assessment and Evaluation" of pupils (and perhaps of the other aspects of the school system) should raise these questions explicitly and consider them at least briefly.

One can get answers to these questions which are specific but "static" in outlook. It is also possible to get answers that, while being specific and operational, have an element of dynamism, fall into a meaningful perspective and, in particular, have the overtones of an "on-going" programme, adapting, assimilating, evolving and actively growing in many dimensions. It will be accepted that the latter type of answer is more useful to the "designer" at all levels than the former.

Hence, "design staff" in the field of "Assessment and Evaluation" may need to consider the scope and nature of an educational system as a whole. An education system - as all "systems" - can be viewed and analysed in many ways. "What particular pattern of analysis is specially relevant?" is an important question.

Usually design staff concerned with "Assessment and Evaluation" are also intimately associated with curriculum design and development. A discussion on "Evaluation and Assessment" must, therefore, be based on a critical review of the general conceptual and methodological aspects of curriculum development.

Curriculum development, viewed in the most general terms, involves the following principal processes:

> Firstly, it involves a complex set of group decisions of a multi-disciplinary nature including many value judgments (the decisions being based on available evidence, studies, expert opinion, body of theory and learning research), the total complex of decisions leading ultimately to a specification of objectives and content for a particular curriculum or educational programmes.
>
> Secondly, having regard to the availability of resources of men, materials and time, resources for learning and instructional procedures are devised to set the conditions under which fruitful learning experiences are made available to pupils. This stage often involves work of an artistic, creative and synthetic nature.
>
> Thirdly, quality control and feed-back of evidence needs to be designed and elaborated to ensure that curriculum plans are effectively realised.

Curriculum development involving the above processes is not a one-stage operation. The third process, indicated above, is essentially the process "Evaluation and Assessment", the valid use of which transforms the entire education programme into a dynamic and continuous one.

Evaluation and assessment are related to pupils, teachers, schools, communities and the whole system. It can also concern itself with the evaluation of simple outcomes in the field of knowledge and facts; or more complicated and more subtle learning outcomes relating to the application of knowledge, analysis, synthesis, etc. Furthermore, it can concern itself with the development of simple and complex skills at a variety of levels of versatility and adaptability; and it can also relate to the evaluation in pupils and teachers of the growth of appreciations, values and attitudes. In any given case one may be concerned with only a limited aspect of evaluation. But the designer and research student may need to see all these possibilities and these may need to be recognised in a coherent perspective. At all times he needs to be conscious of what he leaves out of consideration; that which he leaves out he should leave out deliberately.

2.0 Evaluation in chemistry - one component of the evaluation of second level general education programme

Second level education programmes, and in particular second level general education programmes, have been the subject of study and critical comment at national and international levels in the immediate past. A classic survey in this context is the "World Survey of Secondary Education" carried out by Unesco in 1961. In this study and in related studies significant trends in second level education have been recognised. In addition to world-wide surveys, regional studies are also available, and one that is directly concerned with the Asian region and relating to second level education is "Perspective of Educational Development - 1960-1980. Draft Asian Model", Unesco, Paris, 1964. Similar documentation and analysis are available for the African region and through OECD for Europe. The current educational scene is characterised by an unprecedented expansion of second level general education. Concurrently with this, many acute problems of articulation of various levels and types of education have arisen. In addition, major questions have been raised as to the suitability of curriculum specifications for this large pupil population with heterogeneous characteristics.

Furthermore, the rapid growth of science and technology have brought in major problems of educational design, especially as "pupil-time" is limited. The curriculum of general education needs to be harmonised with the widening concepts of human culture. This aspect has been particularly underlined in a recent study by Unesco (Final Report - Meeting of Experts on the Curriculum of General Education, Moscow, U.S.S.R., 1968; Unesco, Paris).

Contingent on the above changes which are taking place very rapidly indeed, significant changes are occurring in the methods and content of education as knowledge is accumulating at an accelerating pace and new ranges of skills and mechanical assistance are being offered to the teaching profession.

In response to this and many other stimuli, educational designers in the last decade and those charged with the execution of educational programmes have sought to assimilate the new knowledge and techniques and adapt them to existing school programmes to meet the changing needs. How successful these efforts have been is a question that needs analysis. Curriculum development, curriculum implementation and curriculum evaluation have become major activities within general educational programmes. Significant conceptional and procedural guidelines have been recognised. (Vide Unesco Report on Curriculum of General Education, Moscow, 1968, and also "Unesco Regional Programme for Promoting Educational Research in Asia - Curriculum Research in Asian Countries - Workshop Report", National Institute for Educational Research, Tokyo, 1968.)

One of the most significant areas within the curriculum wherein active development work has taken place in the last decade is in the field of science education at the second level. These development activities have occurred in the developed as well as developing countries. Curriculum development in chemistry and the associated programmes of evaluation in chemistry learning and teaching is but one activity within this extensive and world-wide ferment in second level education.

A programme of evaluation in chemistry must necessarily concern itself with objectives of chemistry teaching and learning. And an analysis of the objective of chemistry learning and teaching will be incomplete without an appropriate comparison with general educational objectives associated with second level general education programmes. Statements at the general level and the statements in specific terms for chemistry had to be consistent even though the level of resolution and idiom of the latter are different. A relevant question, therefore, is, what are the priority general objectives of second level general education programmes? In what direction are these objectives changing substantially or in terms of emphasis? It is not possible to review this issue comprehensively here. It may be sufficient to refer to a recent summary at the international level. The relevant reference is given below.

> Meeting of Experts on Curriculum of General Education (Moscow, U.S.S.R., 16-23 January 1968). Final Report ED/CS/4/11, Unesco, Paris, June 1968.

A consideration of the above study as well as such studies as are reported in -

> Invitational Conference on Testing Problems, October 1967 (Chairman: B.S. Bloom); Educational Testing Service, Princeton, New Jersey, 1968.

and

> A Selected and Annotated Bibliography of Studies concerning a Taxonomy of Educational Objectives - Cognitive Domain, R.C. Cox, N.J. Unks; Learning Research and Development Centre, University of Pittsburgh, 1967.

would indicate that many problems of evaluation in chemistry can be profitably viewed as special cases of a general class or problems in relation to evaluation within second level education.

This is further reinforced by the viewpoints implicit in the Workshop Report - Curriculum Research in Asian Countries - N.I.E.R., Tokyo, already referred to.

3.0 Evaluation in chemistry a major process in an education system

An educational system is a complex, highly organised entity. It involves thousands of schools, tens of thousands of teachers and non-teachers, and may involve several million pupils. All these individuals constitute a large organisation with explicit and implicit aims and purposes.

Evaluation is one of the major processes occurring within this complex system.

In recent years attempts have been made to study and analyse education systems using the concepts and techniques of "systems analysis". An attempt of a general nature in this direction is found in a recent study entitled:

> The World Educational Crisis - A Systems Analysis, Philip Coombs, International Institute of Educational Planning, Paris, 1968.

The framework used in this analysis is general and is indirectly relevant. The recognition of an educational system as a dynamic one with diverse "inputs" and "outputs" and responding to many "internal" and "external" pressures is a necessary aspect of the insight for staff at design and research level.

Two modes of analysis having implicit in it a systems approach directly and indirectly relevant to "Evaluation" are given in the Annexe to the study relating to the curriculum of general education cited above. (Report on Curriculum of General Education - Moscow; Unesco, 1968.)

Curriculum design, curriculum development and curriculum evaluation involve complex modes of communication and execution within an education system. Among other things the adequacy of such communication will ensure that the system will adapt and grow. The dynamics of interaction that are brought into being when active curriculum development and curriculum evaluation programmes are initiated within an education system is very complex. The pattern and range of interaction vary from that which occur at the level of policy decision makers (ministers and directional staff) to the various complex interactions which occur at the level of principals, teachers and pupils within schools and classrooms. This latter group forms the primary focus of attention around which the entire education system operates.

Curriculum development and curriculum evaluation need to be designed and consciously directed recognizing the multiplicity of these interactions and with the specific objectives of winning active participation and involvement of all the varying participant groups.

The recognition of the totality of these interactions, and thereafter recognising evaluation in perspective within the system, may require techniques of analysis other than "verbal" ones. Two schematic models which complement one another are given in the Annexe to the Final Report of the Moscow Conference on General Education.

Detailed descriptions of these theoretical constructs are given in the Workshop Report of the N.I.E.R., Tokyo, and in the Ministry of Education, Ceylon, pamphlet "Structural and Functional Aspects of an Educational System - Ceylon, 1967".

In the implementation of curriculum programmes many people are involved. These participants operate at a variety of levels. The participants fall into the following broad categories*:

(1) One group of participants essentially operate at "survey" tasks and analyse the present curricula and teaching programmes, describing the present situation. An attempt is made to isolate significant problems associated with the system.

(2) A second group of participants could be involved in assimilating information such as obtained above and recognising the problems specifically, breaking them into detailed problems of an operational nature and then trying to work out alternative lines of approach towards solution.

(3) A third group of participants could be policy direction and design staff in education systems who seek to assimilate and adapt the information such as that obtained by the above participants, and to plan out teaching programmes for the various education systems in which they function. These participants will be essentially involved in designing plans and implementation programmes using selected preferred lines of approach.

(4) Another group of participants would essentially be charged with tasks such as communicating the plans, with programmes of teaching, to executive levels within the educational systems, leading ultimately to communication and assistance to principals and teachers working in the schools.

---

* Although functionally isolated, the same individuals may thus often be functioning in more than one group.

(5) Still other participants will be engaged in assisting at all these various levels and controlling and supervising the programmes.

(6) The last and most important group of participants could be teachers who will be seeking to implement the curriculum programme in the classroom. This group is central to the whole programme of activity and other functions so as to assist them in their work.

(7) But on-going curriculum programmes, having vitality, must necessarily involve a further critical group of participants whose primary task is the assessment and evaluation of the achievement of curriculum objectives. These curriculum objectives and hence evaluation may relate to aspects of functions of design level staff, teaching level staff and assessment and evaluation of pupil growth and development.

As has been indicated, a curriculum development and implementation programme can fail to operate or, when it operates, fail to operate adequately and achieve desired goals effectively, on account of inadequacies at one or more of these various levels of participation. In short, general education programmes may fail because design staff did not participate adequately with proper consultation and appropriate contacts with schools. At another level the plan and programme staff may have participated in a manner which was not comprehensive, consistent and workable. Again, even when the design is adequate and the plans and programmes are comprehensive, consistent and workable, the message given to the operational levels in the schools and in the classrooms may be inadequate, unclear, ambiguous (full of noise). Curriculum implementation programmes may also fail because of the inappropriate allocation of resources of men, material and time. Even if all the above are adequate, still the quality of implementation may not be high, because supervision and assistance has not been given in full measure. When all of the aspects are perfect, even then the entire situation can be "static" and lacking in progressive characteristics (tending to fossilise) after one stage of development if evaluation programmes are inadequate. Evaluation and assessment are not only essential to ensure an on-going programme, but must occur at all the levels ranging all the way to classroom implementation. Evaluation must be directed towards locating all the various types of inadequacies that have been referred to above. The brief account given would indicate the critical role that assessment and evaluation play in an educational programme. When the pace of change was not very rapid and the sense of urgency relatively low, then the need for feed-back information was less acute than now. With a very rapid pace of change and an acute sense of awareness and urgency, rapid feed-back and sensitive response to evaluation data becomes a vital necessity in any programme of work.

In a programme such as design and implementation of general education programmes where "single-valued real" solutions cannot be worked out from established theoretical frameworks, there is a need to monitor continuously the validity and workability of design programmes at the implementation level. Evaluation is not an activity that is limited to simple evaluation of pupil growth per se. It is true that ultimately the final test of the adequacy of an education system will be the assessment of pupil growth. If this occurs adequately and effectively, then to that extent it may be possible to argue that all other components are functioning reasonably adequately.

Usually, unqualified success of design attempts does not attend a complex implementation programme, such as in the field of general education. Only partial success is achieved. The limited success may be the result of limitations and inadequacies occurring at all the various points in the implementation sequence outlined above. Evaluation must be directed towards achieving an understanding of performance characteristics and also towards diagnosing the particular problems which could arise and cause the partial achievement at any given level.

Curriculum development involving all the above complex of processes is not a one-stage operation. The continuous nature of curriculum development and implementation has to be recognised all the while by all the various participant staff indicated.

Evaluation is an integral part and process within the larger programmes of curriculum development in an educational system. Well planned evaluation programmes can be an immense asset in curriculum implementation.

Ultimately the success or failure of an education programme must necessarily be assessed by the adequacy with which it discharges the assignment given to it by the community it serves. The assessment of community growth itself must therefore become an integral part of the evaluation. The assessment of the pupils, teachers, principals, supervisors, administrators, etc., may well be elements which will make up this final assessment and evaluation.

The school system is essentially an agency established by the community and is the main agency through which a community develops its ideals and attitudes and culture through generations - conserving it, developing it and extending it. Every moment, everyone is actively going through processes of education and change of one type or another. This education may be imparted in an informal manner through the various mass media techniques or in a formal manner through the established institutions such as schools, etc. The focus of attention in the current paper is essentially on the role and function of evaluation within and related to the formal agencies.

At each of the various functioning levels within an educational system, evaluation needs to be carried out in significant dimensions. An evaluation programme to be comprehensive, consistent and adequate in practice must concern itself with all the various possible dimensions, namely:

the dimensions of cognitive achievement*;

the dimensions of psychomotor execution*; and

the dimensions of affective involvement*.

The evaluation of any individual or group at any one of the levels within education must ideally be concerned with evaluating in all of these various dimensions. It may well be that the particular problem in hand and the limitations of current conceptual and technical knowledge and skills may compel one to limit severely the aspects which will be subjected to evaluation. While this may be so it is necessary that design staff should recognise explicitly, and at each level, in each situation, the various aspects which their evaluation programmes leave out of consideration. It would be valid to argue that design staff have no justification for leaving things out by default, forgetfulness and lack of attention. On the other hand, if limitation in knowledge, skills, etc., compel one to leave things out then it would be appropriate to leave aspects out explicitly and deliberately, recognising that all such omissions bring in limitations in relation to the interpretative value and usefulness of the data.

---

* The details of these various dimensions and the sub-levels within them have been described in the context of curriculum development and evaluation in an associated working paper entitled "Theoretical Constructs Relevant in Curriculum Development and Evaluation". This study uses a conceptual analysis which is essentially an adaptation of the analysis of Bloom, Krathwohl et al. in this field. The scheme of analysis in the reference cited is comprehensive, in that it includes the psychomotor domain as well.

2. Evaluation - a Critical Stimulus to Science Curriculum Development in Ceylon

by J. Ratnaike

Generalised patterns of project planning and management, applicable to such diverse areas of human activity as industry, government, educational planning, defence and warfare, have appeared in the last few years. Regardless of the sophistication of the projects, their specific or general nature, or whether they be limited or extensive, the framework of essential, integrated, cyclic, unit operations (decision making, planning, communicating, executing, controlling, evaluation) has been found to be versatile and effective as a conceptual tool for programme design and management.

Useful transfer of these ideas to the planning of science curriculum development has been made. The flow sheet of the operation elements of the Ceylon Curriculum Development Project (Figure 1), since 1958, reflects the essential components and patterns of such generalised analyses.

The purpose of this paper is to summarise the contribution which one important element - evaluation - has made to the Ceylon Project, and to identify areas where further action-research is required. An underlying feature of the action-research programme involving evaluation, and in fact involving all the other elements of the project, is the extent of realistic compromise from ideal situations that was made, particularly because of resource constraints, in a developing country like Ceylon, engaged in a major curriculum development programme. Nevertheless it will also be recognised that even without brilliant and sophisticated break-throughs in ideas and techniques, if a comprehensive, consistent and workable plan is designed, a considerable amount of constructive and productive preliminary activity may be successfully completed - within the restrictive framework of limited time, and human and material resources.

Phase One

Planning and decision making for a project requires data relevant to the situation. Some subjective evidence was available that the teaching in Ceylon was didactic and academic, and coaching for examinations the prime objective of the "teaching". The question naturally arose as to what the characteristics of these examinations were. This generated

FLOW CHART SHOWING MAJOR ELEMENTS OF THE CEYLON SCIENCE/MATHS PROJECT

(from the SEAMES Report on the Ceylon Project)

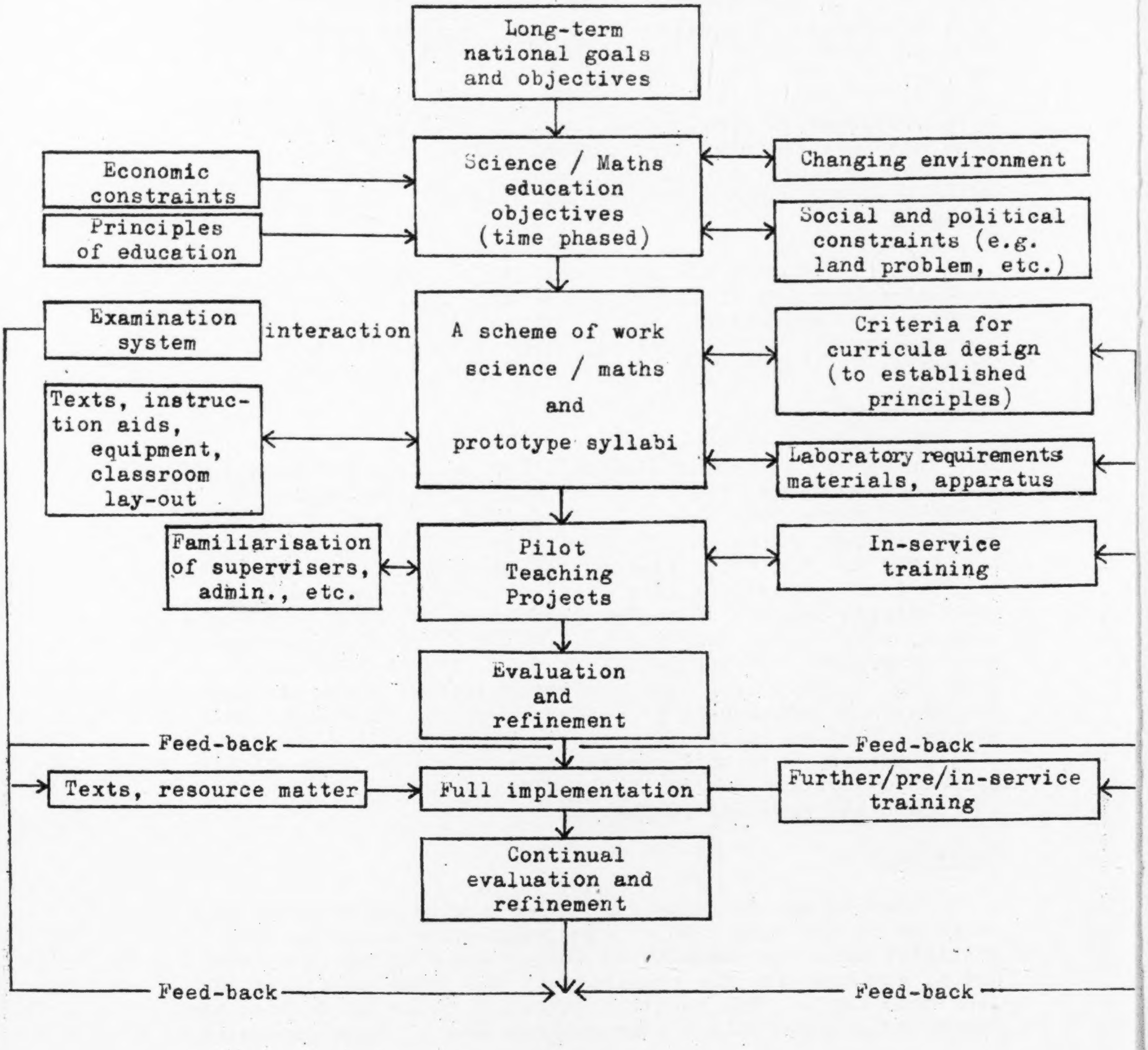

Long-term national goals and objectives
Economic constraints
Principles of education
Science / Maths education objectives (time phased)
Changing environment
Social and political constraints (e.g. land problem, etc.)
Examination system
interaction
A scheme of work science / maths and prototype syllabi
Criteria for curricula design (to established principles)
Texts, instruction aids, equipment, classroom lay-out
Laboratory requirements materials, apparatus
Familiarisation of supervisers, admin., etc.
Pilot Teaching Projects
In-service training
Evaluation and refinement
Feed-back
Feed-back
Texts, resource matter
Full implementation
Further/pre/in-service training
Continual evaluation and refinement
Feed-back
Feed-back

Figure 1

the need to define specifically what objectives were to be examined. Neither the Education Department, which bore the responsibility of teaching pupils for the SSG/GCE 'O' Examination (pupil age 16+), nor the Examinations Department, which evaluated the pupils and issued them certificates, had specified objectives for the examination in significant dimensions sufficient to locate inadequacies in the examination and show guidelines for correction.

The definition of objectives was indeed a major operation, for it meant not only the translation of socio-political objectives into objectives of science teaching, but also defining the objectives in such a manner as to provide operational cues for the design and assessment of science teaching. The pregnant social changes that had taken place in Ceylon since 1956 brought in their wake major changes in the nature, composition, hopes, aspirations and functions of the school population. The percentage of pupils in the age range 14-16 who had some opportunity of utilising their science learnings at higher levels was less than 5%. Of all the 14 year old pupils in the science classes, 95% were to join the general public of Ceylon for whom this was a terminal course in science.

The urgency of the task of curriculum development, and the limited number of trained personnel, prevented an extensive and intensive analysis of objectives in all their multiple dimensions, at the start of the Project. The action-research programme isolated some of the critical ones, in one dimension, and used these for the analysis of the examinations and as a preliminary working base for further work or classifying objectives and planning.

The preliminary set of cognitive domain objectives used was as follows (each of these was further broken down into sub-objectives):

I Ability to recall important information.

II Ability to apply principles qualitatively.

III Ability to apply principles quantitatively.

IV Ability to use the methods of the scientist.

It was recognised that, at this preliminary stage, no attempt would be made to work out or utilise objectives for the affective and psychomotor domains, and at that time (1958) no analyses or taxonomies for these domains were available to the Project.

The analysis was made in tabular form, and the results for the years 1952 to 1958 clearly indicated that factual recall was indeed the major objective being tested in each of the science subjects. The application of knowledge appeared in less than 10% of the questions, and not a single question required the pupils to analyse an experiment and its results for validity, or design considerations, or assumptions, nor to synthesize an experiment for a given situation. No questions appeared in which a pupil had to evaluate a situation in terms of internal criteria, or show his level of achievement in using the nature of proof. The relative crudeness of the objectives framework used did leave some room for argument as to the precise objective under which the questions were classified. Also here was an inherent ambiguity since it was always possible that what was classified as "application" could be "factual recall", for it depended on the vaguely defined learning experiences of the pupils. Nevertheless, it was clear that these examinations required the pupil to exhibit his powers of memory much more than his level of understanding and critical thinking. It was not surprising that when examinations demanded essentially recall of fact, the teaching, geared to the examination, also encouraged regurgitation of fact.

The second characteristic of the examination investigated was content coverage. This was done for the same years as above, and the content areas were defined by the Examination Syllabuses of the Science Subjects, issued by the Examinations Department. The sub-division of the content areas was based on the usual content units as they appeared in textbooks and which were traditionally used in classroom teaching. In tabular form it was clearly visible that certain content areas in each science subject were sampled many more times than others (nitrogen in the chemistry examinations 24 times, equilibrium 5 times, while heat and chemical change not once, in 14 examinations). Furthermore, with a pass score of 35%, and with optional questions present in the science examination papers, a pass could be obtained by a pupil who had covered a minor portion of the content areas in a particular subject. Repetition of stereotyped questions from a relatively limited number of content areas enhanced the possibility of guessing examination questions, and encouraged pupils to leave out large areas of the syllabus.

A further significant characteristic was investigated - the pupil performance in these examinations. In spite of the paucity of content coverage, the fairly high predictability of the examination questions, and the demand for essentially recall of fact at the expense of higher levels of achievement, the number of failures at these examinations was high. Further, those who failed, failed very badly. More than a third of the pupils who sat scored less than 24/100.

This analysis raised the need to investigate the reasons for such failures and wastage, at least in terms of standards and facilities for teaching the curriculum and its relevance, its adaptation to the needs, interests and individual differences of the pupils, and its communication, the amount and quantity of supervision and guidance available to teachers and pupils, promotion policies, and socio-economic factors.

These relatively simple analyses, involving very little training and expense, were able to provide a very useful though limited profile of the examinations and, through them, the kind of teaching that had prevailed up till then. They were also able to isolate several critical problems that needed immediate investigation and solution.

Phase Two

Accepting the fact that examinations exerted great power over the methods of classroom teaching, it was decided that a more valid examination should be held, to start with in chemistry, while still using the existing syllabus. The strategy was to accumulate experience over time, through a series of relatively small changes, and do so in a manner that ensured a built-in mechanism of improvement.

This was done in 1961, and was continued until the revised syllabuses were published. These examinations were specifically designed to correct the defects identified by the previous analyses. To obtain wider content coverage, although the same number of questions as previously was retained, each question was broken into several sub-questions or items. A detailed table of specification was used in the design of the question paper. Since the papers were of the free response type, to enhance the reliability of the marking, the questions in the paper were structured, so that answers had to be specific and short.

In addition, a refined, explicit and detailed marking scheme was produced in co-operation with Chief Examiners (who supervised the Marking Examiners) and the Marking Examiners.

Each Marking Examiner filled in a special data form indicating the pupil performance of the first 40 papers he corrected, and of the last 40 papers he marked. The selection of scripts was thereby assumed to be random and the sample was about 25% of those sitting for the examination. The selection was also assumed to be stratified in the following respects. At the start of the marking, the Marking Examiner would not be very familiar and conversant with the detailed marking scheme, but he would not be bored, or working under the stress of proximty to the marking deadline. On the other hand, when he had almost completed his set of scripts, he could be bored and under stress but would be thoroughly familiar with the marking schemes. The pupil achievement profile of this sample was close to the profile of the total population sitting for the examination.

The data for the performance of each part of each question were collected and analysed for the following characteristics, for all three media, Sinhala, Tamil and English:

(i) Frequency of choice per question and per item.

(ii) Frequency of zero marks per item.

(iii) Difficulty indices per question and per item.

(iv) Ordinary pass in whole paper vs choice of item.

(v) Credit pass in whole paper vs choice of item.

(vi) Ordinary pass in whole paper vs 50% score or more in item.

(vii) Credit pass in whole paper vs 50% score or more in item.

(viii) The distribution across the score range.

((iv) to (iii) were indices of discrimination.) This analysis provided a more detailed profile than before.

Equally important was the fact that all the mechanical aspects of the analyses could be handled by staff with hardly any specialised training or knowledge of statistics. Nor was specialised equipment required. In fact the entire operation for the three years 1961 to 1963 was done without even calculating machines, and with only the assistance of teachers and pupils at the school where trial teaching of the new curriculum was being conducted. Yet it was possible to obtain a variety and detail of data useful as feed-back for curriculum development. These operations served as critical explorations in the techniques in evaluation to be used later, after the new examinations and the new syllabuses had been introduced in 1963. They were also effective for the in-service training for staff in a real rather than a contrived action-research situation.

Phase Three

Having gathered some experience in the use of tables of specifications, and in the preliminary analyses of data, in 1963 the new examination syllabuses were produced, together with prototype question papers. The tables of specifications utilised Bloom's Classification of the Cognitive Domain. For the first time in a national science examination, an outline of the objectives (other than content to be tested) was also included in the syllabus.

Each science examination now consists of a multiple choice (objective) type paper, and a free response (essay) type paper, with a relative marks weightage of 2:3. Also an extensive and detailed programme for further action-research in evaluation has been drawn up, and used for the new examinations since 1965. By 1965 a number of personnel had already been trained, and superior facilities, such as computer-time and calculating machines, were available.

The study of the examinations is being done on a 5-10% national sample, stratified in terms of educational regions, type of school, medium, sex. The kinds of analyses involve performance correlations among the objective type papers and the essay type papers within each science subject, and across the science subjects, correlations made across first language, mathematics and second language, for validity studies; difficulty and discrimination indices for item analysis; and effects on pupil performance due to factors such as sex of pupils, geographical region, type of school, urban/rural, teacher experience and qualification, parental education, occupation and income. As a basis for investigating predictive validity of the examination, a follow-up study of each sample will be carried through for ten years.

The production of a general aptitude test has started, and trial testing will begin in a month's time. This will be used later as an additional dimension in the validity studies.

While the investigations mentioned above were of problems having direct relevance in the present stage of development of the theory and practice of measurement and evaluation, and capable of being wholly or partially solved on a short-term basis, the Project was not unmindful of problems which, though significant, were of such a nature that they required long-term fundamental research and a significant break-through in the conceptual viewpoint for their solution. The working papers produced on the restructuring of the Bloom classification, and the outline analysis of the psychomotor domain are an example of this. Other problems are those which arise as a result of the assumptions, limitations and restrictions that apply to widely used measures, theories and methods of education measurement and interpretation, such as the nature and characteristics of:

the process of measurement,

the process being measured,

the instruments used for measurement,

the interaction between one or more of the above.

The limitations

The cyclic operations (decision making, planning, communicating, executing, controlling, evaluating, ideally take place at each level of policy making and/or execution. In the field of science curriculum development, one level at which these cyclic operations take place may correspond to the community and its functional organs, such as Parliament and the Ministry of Education.

The next level may correspond to the Directorate of the Ministry of Education; the third level to field and supervisory staff; the fourth to the principals of schools; the fifth to the teacher, and the sixth to the pupil. Each level would contain the operations of decision making, planning, communicating, executing, controlling and evaluating, and each level would be articulated with the next through the operation of communication.

In this context, for the teacher and pupil levels, the widely used triangular pattern (Figure 2) may be thought of as a simplified and reduced model of the general structure (Figure 3), where decisions in relation to content of teaching, planning of the organisation of this content, the mechanics of communication and the controlling of the pupil are coalesced into a general activity described as "Teaching". "Evaluation" is considered apart. This particular reduced pattern is helpful in focussing attention on evaluation. (In considering contexts other than evaluation, where the focus of attention may be another aspect of the teacher-pupil situation, another reduced form may be more appropriate. The reduced form also depicts the continuously changing and repetitive role of the teacher in the classroom, as one who operates to facilitate learning, as one who assesses pupils' growth.)

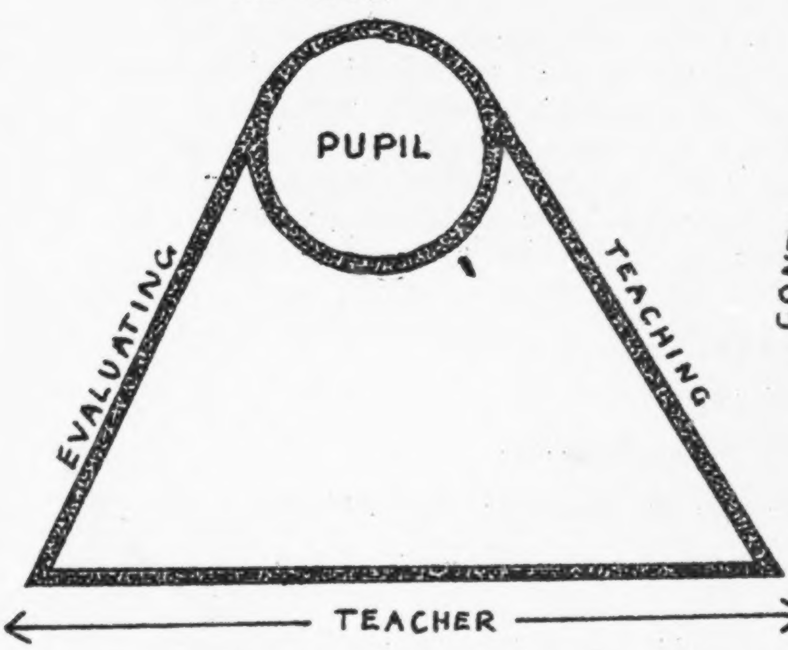


Figure 2

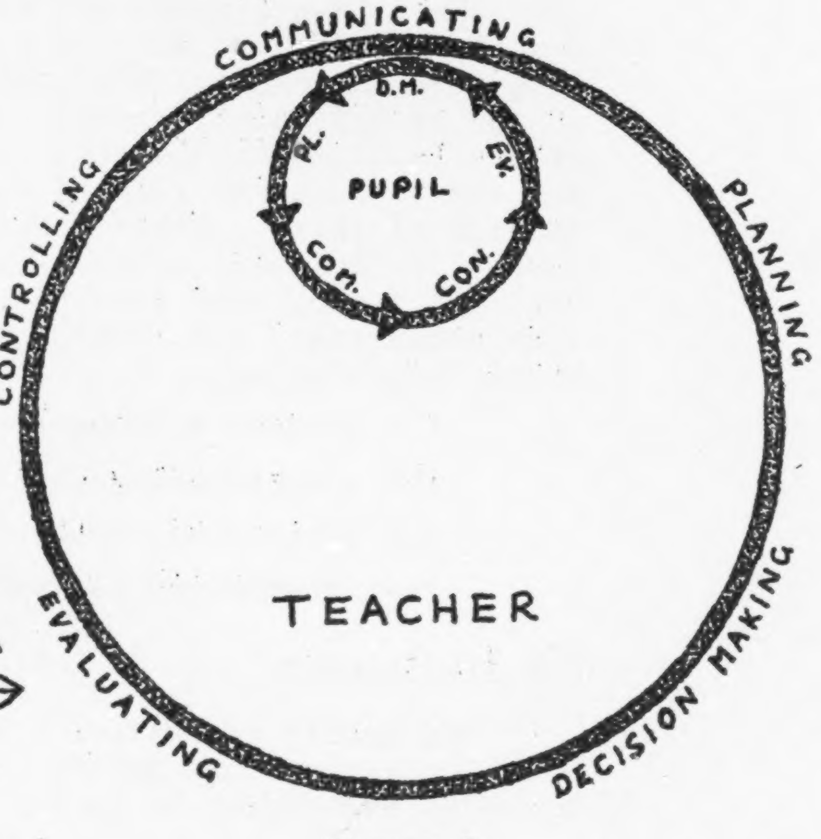


Figure 3

At each of these respective levels of execution, and for each unit operation, within each level, the process of evaluation has to be considered as an inseparable component and an essential means of feed-back and self-correction.

While the problems of pupil achievement evaluation have been investigated fairly extensively, at least in the cognitive domain, because examinations and examination results of pupils have an immediate and tangible socio-economic importance and high powers of motivation, and even though their contribution by way of feed-back of pupil performance and specification of the test items is very significant and effective, it must be remembered that measurement of the change in behaviour of 16 year old pupils is not the end product of this series of operations which started with the conversion and translation of socio-political objectives into teaching objectives and filtered through the various levels of execution.

The end product is the type of behaviour exhibited by the pupils after they leave school and become citizens of the country, whatever their fields of endeavour may be, as buddhist monks or teachers, betel sellers or bus conductors, doctors or politions,**lawyers** or housewives, engineers or farmers. A pupil's achievement at 16+ is not necessarily an indicator of such an end product, and it is certainly not a direct indicator. Ideally the process of end product evaluation should involve the evaluation of the magnitude and direction of change in the community. Even if it is temporarily assumed (however incorrectly) that pupil growth and achievement at 16+ may be taken as the end product, the indices of achievement and growth of pupils (as represented by examination results are once, twice or thrice removed from many of the major elements of the programme, such as decision making, communication, controlling. These indices are products of highly complex processes as of interaction, both positive and negative, among all the operations of all the levels. To evaluate these separate operations by extrapolation from examination results of pupils alone is indeed a hazardoux process, and one unlikely to provide signals of sufficient specificity and discrimination to be used for improvement of the major operations.

The Ceylon Project recognises this very serious drawback, and, while exploiting the examination achievement data of pupils to the maximum, has initiated action-research programmes to develop finally means of direct evaluation of all the operations of science curriculum development, at all levels of execution.

3. Improvement of Evaluation Practices

A study of the G.C.E. 'O' science examination

by P. Samaranayaka

The G.C.E. 'O' examination in physics, chemistry and biology based on the revised syllabuses was first held in December 1965. The study of the examination and pupil performance was done on a national stratified random sample selected prior to the examination.

Nature of the sample

Of the pupils appearing for the examination, only those appearing for the first time from government and director managed schools and offering all three science subjects (physics, chemistry and biology) were considered for the sample.

The sample strength was about 10% of the above population and was selected in terms of the following strata:

(i) Educational region.
(ii) Medium of instruction.
(iii) Grade of school.

In Ceylon there are fifteen educational regions broken down in terms of their geographical location. There are two media of instruction: Sinhala and Tamil. For the purpose of this study the schools were classified into two groups: schools terminating at the G.C.E. 'O' science level and the G.C.E. 'A' science level.

The population was broken down to clusters where a cluster represented a school unit.

Operations in the sample selection

On the basis of information received from the schools regarding their G.C.E. 'O' entries for the December 1965 examination, a list of schools containing the following particulars was prepared.

(i) Name of school.
(ii) Educational region to which the school belonged.
(iii) Media of instruction of the G.C.E. 'O' science streams.
(iv) Type of school (whether G.C.E. 'O' terminal or G.C.E. 'A' terminal).

(v) Type of local body to which the school belonged (whether Municipal Council (MC), Urban Council (UC), Town Council (TC), Village Committee (VC)).

(vi) Number of pupils entered for the examination in each of the media, Sinhala and Tamil.

(vii) Pupil enrolment in the school from Grade VI upwards.

Subject to the stratification referred to earlier, three different samples were drawn at random. These three samples were examined by a panel consisting of the Deputy Director General (Secondary Education), the Chief Education Officer (Science) and another Senior Chief Education Officer with wide experience in the field. Their judgment was sought to select the sample which was most representative in terms of an urban/rural variable (whether pupil's home was in a rural area) which was not a strata factor considered in the selection of the three samples. The urban/rural background was used as a crude index to reflect the socio-economic background of the pupil.

On this basis 42 schools (clusters), of which 28 presented candidates in the Sinhalese medium while the other 14 in the Tamil medium, were selected. The total number of pupils in the Sinhala medium amounted to 688 and in the Tamil medium 393, making a total of 1081 pupils. The sample was called the National Stratified Random Sample.

Procedures for correction of scripts of the sample

A specially selected panel of examiners were entrusted with the correction of scripts in physics, chemistry and biology of this sample. The panel consisting of 14 examiners per subject worked under the direct supervision of the Controlling Chief Examiner of the respective subject, for a period of ten days. In addition to the correction of scripts, these examiners were called upon to fill in a sheet (Data Sheet I) in duplicate in respect of each of the pupils in the sample. The pupil response for each of the 40 questions in paper I (irrespective of whether the response was correct or incorrect) was entered in Data Sheet I. In addition the score per question of paper II was also entered in this Data Sheet I.

Finally a summary of the data in Data Sheet I was made. This summary gave the following particulars in respect of each cluster. The number of pupils selecting each choice of each question in paper I, the number of pupils who omitted a particular question in paper I and paper II, the number of pupils who failed to score in each of the questions in paper II. Data Sheet I, in duplicate, had to be completed by these examiners under the direct supervision of the Controlling Chief Examiner of the respective subject.

Mode of analysis of Data Sheet I

Data Sheets I were scrutinised across subject, and any pupil who had absented himself from any of the subjects physics, chemistry or biology was struck off from the sample. The number struck off from the sample on this basis amounted to 29, which was less than 3% of the sample. The final sample had a total population of 1052 pupils.

Data Sheet I was used to process the following information.

(i) Difficulty index of each of the questions of paper I, in respect of each medium and total sample.

(ii) Response per cent in respect of each of the choices of the questions of paper I in respect of each medium and total sample.

(iii) Percentage of omissions in respect of each of the questions of paper I in respect of each medium and total sample.

(iv) Discrimination coefficients of each of the questions of paper I.

Other data

A file is maintained in respect of each of the pupils in the sample which contains his G.C.E. 'O' answer scripts of physics, chemistry and biology. In addition a copy of his cumulative record is maintained in his file. His performance through school and any other educational institute is also kept track of and these particulars are also recorded.

In March 1966 the Vedanayagam Non-Verbal Intelligence Test and the Jayasooriya Verbal Intelligence Test were administered to these sample pupils and these work sheets are also kept in their respective files.

In addition to Data Sheet I, another sheet for recording pupil data was designed. This sheet, referred to as Data Sheet II, contains the marks of the pupils in the sample in the following papers.

(i) First language
(ii) Second language
(iii) Mathematics / arithmetic
(iv) Physics
(v) Chemistry
(vi) Biology
(vii) Vedanayagam Non-Verbal Intelligence Test
(viii) Jayasooriya Verbal Intelligence Test

These marks are recorded in Data Sheet II in terms of clusters. In addition to these marks, Data Sheet II records the teacher qualification and experience in respect of physics, chemistry and biology.

Teacher classification is in terms of five broad groups:

(i) Graduate with over 5 years' teaching experience.
(ii) Trained with over 5 years' teaching experience after training.
(iii) Graduate with less than 5 years' teaching experience.
(iv) Trained with less than 5 years' teaching experience after training.
(v) Others.

A record of the pupil's socio-economic status is also maintained in Data Sheet II. These are maintained in respect of:

(i) parental income;
(ii) parental occupation;
(iii) parental education;
(iv) type of local body in which the pupil has been residing for the longest period of time.

Mode of analysis of Data Sheet II

Analysis of Data Sheet II is classified into two main groups:

(i) Calculation of strata population, mean, median and standard deviation of physics I, physics II, chemistry I, chemistry II, biology I, biology II, first language, second language and mathematics/arithmetic in terms of each of the strata on which the sample was selected.
(ii) Calculation of the inter-correlation among papers and different combinations of papers for the total sample.

Analysis of other data

Data in terms of each pupil's progress through a minimum of 5 years after sitting the G.C.E. 'O' examination is in the process of being collected. Analysis and interpretation of this data has just commenced.

1966 G.C.E. 'O'

The nature of the population for the 1966 G.C.E. 'O' examination was similar to that of 1965. However, the size of the sample was brought down to about 6%. This step was taken with a view to limiting the sample population to handlable numbers as it was predicted that with time the population would increase significantly.

The nature of the sample for 1966 was slightly modified from that of 1965 in that sex was also introduced as a stratum in addition to educational region, medium of instruction and the grade of school.

The breakdown of the population into smaller units (clusters) where a cluster represented a school unit was found to be rather unsatisfactory as certain schools had as many as two hundred candidates who were eligible for selection to the sample. If this type of cluster was selected for the sample it was evident that the sample would be heavily loaded in one direction or another. However, to eliminate such a school on these grounds would necessarily make the sample less representative. As a compromise solution it was decided to break down each school unit into simpler clusters where the maximum pupil strength per cluster was fixed at 30. On this basis a school which has 100 candidates eligible for selection to the sample is broken down to 4 clusters where each of the first 3 clusters has 30 pupils and the last carries the balance of 10 pupils.

Mode of selection of the sample

As opposed to the 1965 procedure where the basic information regarding G.C.E. 'O' entires were collected from the schools, in 1966 this information was collected directly from the G.C.E. 'O' entry lists of the Department of Examinations.

In this manner the following information was recorded.

(i) Educational region.
(ii) Name of school.
(iii) Grade of school.
(iv) Type of management of school.
(v) Medium of instruction.
(vi) Number of candidates in each sex.
(vii) Centre number.
(viii) Inclusive index numbers of the candidates.

The information collected was condensed and recorded educational region wise. With the help of this information the total strength of the sample and the strength per stratum were determined. Having numbered the clusters in the serial order, five different samples of clusters were selected at random subject to the above stratification using a table of random numbers. From these five samples one sample was selected by a panel of three members consisting of the Deputy Director General of Education and two others so that this sample would reflect a balance in terms of the socio-economic factors that operate in the population. This selection was done on similar criteria as used for the final sample selection of 1965. As the Sinhala population in the samples amounted to approximately 13% of the population, only every other candidate of the Sinhala sample was included, while the Tamil sample was taken in toto. On this basis the maximum strength per cluster of the Sinhala medium population was brought down to 15 while that of the Tamil medium population was 30.

The sample for 1966 consisted of 90 clusters representing 88 schools and the total pupil strength was 802. Of these, 73 clusters consisting of 552 pupils were in the Sinhala population sample and 17 clusters consisting of 250 pupils were in the Tamil medium population.

Panel of examiners

The panel of examiners for correction of scripts consisted of 18 for each of the subjects physics, chemistry and biology. As in 1965, they were entrusted with both marking of scripts as well as entering of Data Sheets I under the supervision of the Controlling Chief Examiners of the respective subjects.

Other data

The collection of other data of the sample was done in a manner similar to that of 1965.

Mode of analysis of data

Analysis of Data Sheets I was similar to that of 1965. Nineteen pupils from the Sinhala population and 10 pupils from the Tamil population had to be deleted from the sample, as pupils had absented themselves from one or more of the papers in physics, chemistry and biology. The sample was reduced to 773 pupils as a result of these deletions. Difficulty indices, response per cent and discrimination coefficients were worked out for each of the items in paper I of physics, chemistry and biology.

The analysis of Data Sheet II was slightly modified from the procedure adopted in 1965. Analysis of scores of the Vedanayagam Non-Verbal Intelligence Test and the Jayasooriya Verbal Intelligence Test was not attempted, as this G.C.E. 'O' population was outside the range for which these tests had originally been designed. Analysis of scores of the total sample as well as per stratum was attempted in terms of the following papers:

(i) Physics I
(ii) Physics II
(iii) Chemistry I

(iv) Chemistry II
(v) Biology I
(vi) Biology II
(vii) Specific combinations of the above six papers
(viii) First language
(ix) Second language
(x) Mathematics

In this analysis the population strength, mean, median and the standard deviation were determined in terms of each of the strata.

In addition, coefficients of correlation among papers and of certain combinations of papers were attempted.

Other data

The performance of this sample, too, is to be followed through the next few years for ultimate analysis. All the pupils in the sample are kept track of and a file is maintained in respect of each of these pupils.

1967 G.C.E. 'O'

The nature of the population for the 1967 G.C.E. 'O' examination was similar to that of 1965 and 1966. The size of the sample was brought down further to 5%.

The nature of the sample for 1967 was similar to that of 1966 and the same strata factors were considered for the selection of the sample. The breakdown of the population into clusters was slightly modified in that the maximum strength of a cluster was fixed at 15. This was done by breaking down the school unit to simpler units of maximum strength 30 and then having selected these units of maximum strength 30 for the sample at random, every other candidate in these units was considered for the clusters of the final sample.

Mode of selection of the sample

The selection of the sample was again done on the basis of information collected from the entry lists of the Department of Examinations. On the above basis 5 different samples of clusters were selected at random and a panel of three members inclusive of the Deputy Director General of Education selected one of these so that the sample would reflect a balance in terms of the socio-economic factors that may operate in the sample.

The sample for 1967 consisted of 100 clusters representing 97 schools and the total sample strength was 779. Of these, 72 clusters consisting of 589 were in the Sinhala medium population and 28 clusters consisting of 190 were in the Tamil medium population.

Panel of examiners

The panel of examiners for correction of scripts consisted of 20 for each of the subjects physics, chemistry and biology. As in 1965 and 1966 they were entrusted with both marking of scripts as well as entering of Data Sheet I under the supervision of the Controlling Chief Examiners of the respective subjects.

Other data

The collection of other data pertaining to the sample is currently being carried out. The principals of the respective schools in the 1967 sample have been contacted and the names of the pupils selected for the sample from each of their schools have been communicated to them. The required particulars of these pupils are now coming in.

Mode of analysis of data

Analysis of Data Sheet I was similar to that of 1965 and 1966. 19 pupils from the Sinhala medium population and 5 pupils from the Tamil medium population were deleted from the sample on criteria similar to those adopted in 1965 and 1966. The sample was reduced to 755 pupils as a result of these deletions.

Difficulty indices, response per cent and discrimination coefficients were worked out for each of the items in Paper I of physics, chemistry and biology.

For the 1967 sample only this data has been analysed so far. Analysis of Data Sheet II has not been attempted as the collection of the data has not yet been completed. However, an analysis similar to that of 1966 is envisaged for Data Sheet II.

Interpretation and use of data

The G.C.E. 'O' examination data for 1965 and 1966 have been analysed to an appreciable extent and some tentative interpretations of this processed data have been attempted.

One of the major changes in the examination design that was effected in 1965 was the introduction of an objectively markable fixed response question paper for each of the subjects physics, chemistry and biology. This paper, carrying 40 questions (Paper I) was administered in addition to an essay-type question paper (Paper II). Paper II had the characteristics of the essay-type question papers of the 1961-1964 G.C.E. 'O' chemistry examinations.

At this juncture it is desirable to investigate whether Paper I is testing the same things as Paper II. A way of looking at this may be to examine how the Paper I's of the science subjects correlate with the Paper II's. The correlations for Physics I + Chemistry I + Biology I vs Physics II + Chemistry II + Biology II for 1965 and 1966 are of the order of + 0.9. The correlations of Paper I to II for physics, chemistry and biology taken individually are also of the order of + 0.6 and + 0.7 for 1965 and 1966 respectively. This order of correlations may be interpreted as significantly high. It means that the new design of the examination has helped in the improvement of marking reliability and content and objective validity, while essentially testing similar factors.

The absence of suitable aptitude tests and I.Q. tests for the Ceylon G.C.E. 'O' level child and the absence of horizontal projections usable for predictive purposes is unfortunate. However, the criterion of what is traditionally called a good student, i.e. performance not only in science subjects but also in first language and mathematics, may be used as an index for judging whether the science papers in the examination are behaving significantly differently to the other papers of the G.C.E. 'O' examination. This may be achieved by examining the correlation coefficients between the performance in all science papers (i.e. Physics I & II, Chemistry I & II and Biology I & II) with first language and mathematics. These are of the order of + 0.7 for mathematics and + 0.4 for first language for both 1965 and 1966.

Some tentative trends in achievement are becoming visible in terms of the urban/rural background of the child, social, economic and educational status of the parent. But further data is required in support of these trends before they may be presented for discussion. Similarly predictive validities of the examination would require several more years while cohorts of the stratified random sample occupy positions at the universities, in other higher educational institutes and in the professions.

The data has in part been used for the design and moderation of the science papers at the G.C.E. 'O' examinations of 1966 and 1967. The data carrying an item analysis of the previous examination is made available to the Controlling Chief Examiners and the moderators of physics, chemistry and biology at the time of setting and moderating the G.C.E. 'O' papers.

Some of this data has also been used to isolate specific areas of difficulty for designing in-service programmes.

At this stage no attempt is made to make assertions one way or another. However, with time and further data it is proposed to report other significant trends and problems on the teaching and learning of science in Ceylon at the G.C.E. 'O' level.

4. Notes on Structural and Functional Aspects of an Education System

(Extract from pp. 23-28
of Ministry of Education Ceylon publication)

The Ministry of Education has in the recent past initiated and carried out a programme of work relating to the General Certificate of Education, Ordinary Level, science subjects. In the years immediately preceding 1960 several aspects related to the science teaching situation at the General Certificate of Education, Ordinary Level, were carefully examined and analysed. In particular, the specifications for teaching relating to the General Certificate of Education, Ordinary Level, physics, chemistry and biology were considered. The examination syllabuses at the General Certificate of Education, Ordinary Level, for these subjects were critically analysed. Specific problems and limitations were recognised and decisions were made to pursue development programmes directed towards the improvement of the specifications for teaching and the specifications for examination.

In making these decisions, relevant data from within the educational system were used. In addition, relevant data from other educational systems elsewhere in the world were examined and considered.

In terms of the decision to design improved specifications for science teaching at the General Certificate of Education, Ordinary Level, senior experienced teachers from within the Ceylon educational system, and outstanding consultants from outside, were invited. A team of workers numbering approximately 10 to 20 worked on the programme continuously. At intervals of time, groups, as large as 50 to 100, gathered together, for short periods, to participate in specific aspects relating to design and planning. Professional viewpoints such as are available with scientists, university science teaching staff, and research organisations, such as the Ceylon Association for the Advancement of Science, were channelled to the programme to make the plan adequate. The extent of participation was limited by time and funds.

The communications so designed - in this particular instance "Syllabuses of Instruction and Schemes of Work" and the Teachers' Guide Notes - were not directly issued to the field by the supervisory staff until some executive action had been taken.

Some of these executive operations related to simple routine matters such as writing the syllabuses of instruction and schemes, editing and printing them in sufficient numbers. It was also necessary to set afoot relevant machinery for the storage and distribution of such material. It will be recognised that thousands of copies of schemes of work and syllabuses of instruction for issue to over 500 schools needed specific organisational and administrative action.

A major preliminary executive action was the organising and running of pre-trial teaching of the material in a few selected schools for checking out the time base and other such aspects.

There were also such matters as the design of suitable standard minimum lists of equipment and the ordering of the new materials necessary for teaching the scheme. The recruitment of additional supervisory staff and their training to handle the new schemes were also necessary.

Having pursued these relevant executive matters, and others not mentioned above, the plans and specifications for the teaching (General Certificate of Education, Ordinary Level) of physics, chemistry and biology were communicated to the teachers, principals and other field staff. This entailed a particularly complicated programme of communication to over 500 schools located in many educational districts. Though the syllabuses of instruction contained relatively detailed "Forewords" and "Prefaces" relating to the programme, it was nevertheless necessary to arrange for personal meetings of the design staff with the field staff. This was done in all the districts to a limited extent depending on the availability of time and other resources.

The pupils' performances in the field at the General Certificate of Education, Ordinary Level, examination were carefully analysed in all the subject areas, using a carefully stratified random sample of pupils selected from the examination enrolment. The performance of this sample of pupils was subject to detailed study and analysis.

The whole programme of decision making, planning, communicating, supervision and evaluation is one that proceeds cyclically and unendingly. Therefore as a logical consequence of this recognition there should be continuing action-research programme directed towards assessing pupils' performances at the General Certificate of Education, Ordinary Level, in the relevant subject areas annually. This provision and decision will lead to the availability of feed-back information. In the light of such feed-back information, the total programme could be kept under continuous review and new decisions could be made and new programmes of planning and reconstruction could be set afoot so that the up-graded and the up-dated pattern of teaching and learning would be maintained at a high level of adequacy through time.

The principal of a school who had received the communication that science was to be taught in the General Certificate of Education classes on the basis of modified syllabuses of instruction and on the basis of amended syllabuses of examination needed to ask a number of questions. Some of them were as follows:

How much time was to be given to the subject?

How were these periods to be organised in the day and in the week?

What problems were there in organising the periods for work in the laboratories?

How many double periods needed to be provided?

Where should the theory periods be taught?

Which teachers should be commissioned to do the modified syllabuses?

Is the equipment in the school sufficient for doing the modified programmes?

.... and a series of similar questions.

Having analysed this situation in the school and recognised the major problems in collaborative discussion with the teachers he would ideally plan out a programme of work in general; design the timetable, in particular; and attend to such other matters of planning as were necessary.

Once this stage had been carried out, the teacher would be instructed to teach the modified syllabuses to the pupils in the General Certificate of Education classes.

When the teachers are at work, classroom supervision would need to be provided. The principal may need to give assistance and guidance. This may relate to how the modified programme of work should be approached and worked on. He may in fact be in a position to advise directly on pedagogical and methodological aspects.

Once a teacher continues to work reasonably adequately on the basis of the communication and guidance and supervision provided, the principal may need to associate himself with the teachers and arrange to evaluate the progress of the work. In particular he may need to ask whether the pace of teaching is as recommended; whether the practical work and the manner of doing it are as specified and other similar questions. In addition, principals, with the collaborative assistance of the staff, will need to set afoot the necessary machinery for teachers to design term tests and similar instruments of evaluation of pupil growth.

This will give the principal significant feed-back material which he may find useful for the purpose of reporting on the programme of work.

In such reporting the principal himself taking note of the evaluation of performance of the system may indicate the lines of new decision making and planning which the programme in his school would seem to demand.

The teacher who has picked up the communication could now proceed to study the situation and such study will entail the consideration of such questions as follow:

What are the features of the revised schemes of work?

How are they organised?

How do I use them?

What particular major principles underly these schemes?

Do I have the equipment and materials to teach them?

What particular difficulties will my pupils have in learning in the modified way?

How will I help them? Etc., etc.

Having recognised the nature of the situation, significant problems associated with it and having taken note of the preferred lines of approach, the teacher could proceed to plan his lessons, execute them, assist and supervise the pupils at work and evaluate his and their progress in appropriate ways.

It will be observed that one of the critical features of this schematic model is that each person who picks up a communication proceeds to amplify it and prepare detailed plans appropriate to his level prior to execution and proceeding with the programme.

5. Oral Evaluation, with Particular Relevance to Soviet Schools

by S.A. Balezin

At present every nation is living through a period of a large-scale scientific and technological revolution. The inflow of scientific data is increasing at an unbelievable speed. That is why the school and the school teacher are facing the problem of deciding on what should not be taught rather than on what should be taught to the pupil. In other words, the problem of selecting such teaching material that would deal with the fundamentals of science on a modern level, and at the same time could be within the pupil's comprehension, considering his age-group peculiarities, is becoming increasingly important.

Young men and women starting their independent life will inevitably become immediate participants in technological progress.

In the advanced countries of the world, a considerable increase is being observed in the number of workers with secondary education. Today it is not infrequent to encounter instances where the workers immediately engaged in production become research workers in the full sense of this word. Such workers need to have not only skilful hands but also a clear creative thinking and effective knowledge. This means that the pupils should not only assimilate a certain amount of knowledge but also master, to a certain degree, the methods used to acquire such knowledge, starting with the simplest techniques of research, for instance the qualitative analysis of acids and salts, and up to solving the complex problems arising in production processes.

That is why pupils should be trained in such a way that they could master, to a certain degree, the methods and techniques used in the jobs they will take up after finishing school, to facilitate the solution of the questions on which they will have to take decisions in production and in everyday life. This is possible only on the basis of a certain amount of firm knowledge.

Chemistry plays an outstanding part in the life of the contemporary human community. Therefore, chemistry as an instructional subject in the school curriculum is of special importance.

Forms of current evaluation of the pupil's knowledge: oral evaluation

In oral evaluation, the verification of the pupil's knowledge includes recapitulation and it is one of the teacher's weapons to improve the knowledge not only of the pupil who is answering the teacher's question but also that of the rest of the pupils.

Oral evaluation is widely used in current evaluation and is practised in almost every lesson in chemistry. It serves as both a form of control and instruction, thus improving the pupil's knowledge and his ability to relate it to the listener.

Oral evaluation is made before introduction of new material, during its presentation and at a special lesson devoted to evaluation, discussion and consolidation of the pupil's knowledge.

In the process of oral evaluation, the pupil is asked questions on the given topic covering the material studied. Also he is given problems to solve. In oral evaluation it is possible to establish the amount and depth of the knowledge, its firmness, as well as to find out whether the material has been assimilated systematically and consciously.

These are the guiding rules for preparing questions for oral evaluation:

Questions must take into account the amount of material envisaged in the syllabus, be brief, simple and easy to understand, avoiding too much detail, hints and ambiguities.

Along with the selected particular questions it is also advisable to give the pupils small topics for oral discussion (talks). In class VII, for instance (in Soviet schools chemistry as an individual science discipline is introduced in class VII) this topic can be "Burning sulphur in oxygen"; in class VIII, "Chemical properties of chlorine"; in class IX, "Acetic acid", etc.

The pupil's talk on the given topic enables the teacher to assess not only the pupil's knowledge but also his ability to present the material consistently, systematically and clearly.

Besides this, pupils are invited to solve problems. This serves as an effective means to find out whether the pupil realises the data he has learned.

Normally, individual current evaluation cannot be easily combined with the active work of the whole class. Besides, it takes time.

One of the forms of oral evaluation which can overcome this disadvantage is to call to the blackboard two or sometimes up to four pupils at a time, when each of them is given an individual assignment. While these pupils are thinking over their questions or solving the problems, the teacher can work with the rest of the class. The answers of the pupils who are at the blackboard are discussed with the participation of the entire class.

When pupils answer, it is important for them to be able to use prepared charts, tables and various chemical substances.

Many teachers, parallel with this form of evaluation, use a so-called group test talk with the whole class on the material studied. This talk takes part of the lesson, or sometimes the entire lesson. The teacher puts small questions to the class. The pupils give brief answers without leaving their seats, and the teacher gives marks to those pupils who have been particularly active.

This evaluation/talk is sometimes devoted to summing up the material of a large section of the syllabus. For instance, "Interaction between classes of inorganic compounds or general characteristics of the elements of the halogens group".

Tentative criteria for evaluating the pupil's knowledge

In the U.S.S.R. a five-mark system of evaluation is adopted, the highest mark being the 5. Marks for one term, half year and year are given taking into account the results of current evaluation of the pupil's knowledge and skills. When final marks are given the virtual level of each pupil's knowledge and skills should be considered.

In evaluation of a pupil's oral answer, the mark 5 is given provided the pupil:

(a) shows understanding of the chemical essence of studied phenomena and processes; correctly formulates chemical concepts, laws and theories; knows the properties of the studied substances and can identify them; correctly reads and makes up formulas of substances and conclusions of chemical reactions, gives them correct qualitative and quantitative interpretation;

(b) supports the theory he discusses by concrete examples, knows the practical significance of chemical laws, theories and processes as well as the use of particular substances;

(c) gives a complete and consistant answer displaying independent judgment and assimilation not only of the material being studied but of the previous material as well; when necessary, proceeds from the knowledge learnt when studying other curriculum subjects and applied in the process of labour and vocational education.

The mark 4 is given if the answer, in the main, meets the same requirements established for the mark 5 but the pupil:

(a) makes mistakes in accuracy and insignificant mistakes in relating facts, definition of concepts, laws and theories and in their interpretation while being able to correct all this easily with the help of the teacher; or

(b) gives an an answer only narrating the text of the manual; while in replying to the teacher's additional questions displays due understanding of the material studied.

The mark 3 is given if the pupil shows the knowledge and understanding of the basic instructional material but:

(a) experiences difficulty in supporting, without the teacher's help, the theory being related, by concrete examples, or in explaining the practical significance of chemical laws, theories and processes;

(b) experiences difficulty in summing up the material and drawing conclusions on it, though relates the factual material fully enough;

(c) makes one or two mistakes in reading and writing the formulas of substances and conclusions of chemical reactions, or in their quantitative and qualitative interpretation; or

(d) gives a schematic answer skipping essential details, relating, in the main, the text of the manual while displaying insufficient understanding of the individual tenets that he freely relates.

The mark 2 is given if the pupil:

(a) does not know, or fails to understand, most of the most essential part of the instructional material dealt with in the questions put to him; or

(b) in total, makes mistakes and commits errors indicated in the requirements established for the mark 3; or

(c) even when given the help of the teacher, fails to make use, while answering, of the knowledge assimilated earlier in the course of chemistry and in other curriculum subjects as well as the knowledge and data accumulated in the process of labour and vocational education.

The mark 1 is given if the pupil shows complete ignorance of, and fails to understand, the basic curriculum materials.

Conclusion

In Soviet schools, everyday current evaluation of pupils' knowledge and skills forms the basis of evaluation of pupils' knowledge. Use is made of oral individual evaluation, written tests, experimental tests and experimental problems. However, research that has been and is being carried out in the U.S.S.R. shows that further work is needed to improve the methods of verification and evaluation of pupils' knowledge and to achieve a better control over the knowledge and skills of pupils as well as to achieve more profound knowledge and skills.

6. Report on Evaluation Workshop in Chemistry Peradeniya, Ceylon, 2-12 August 1968

by J.C. Mathews and J. Ratnaike

Aims of Workshop and this report

The Workshop was part of the in-service training of Ceylon teachers and other education staff in the construction of examinations in chemistry at the Ceylon General Certificate of Education, Ordinary Level. The long-term objective is the production of a team of question writers and examination constructors who would assist in the design of 'O' level examinations and act as nuclei for further training of teachers.

This report sets out the sequence of operations during the Workshop and the materials which were the outcome of the Workshop. In so doing, it is hoped that the report will serve as a guide to other countries who wish to conduct similar workshops. The scheme of work is not set out as an ideal one; it was the first comprehensive workshop to be undertaken in this country and those in the future may well be modified in the light of this experience; nor are the questions which were constructed of equal merit. Nevertheless, a consideration of both the faults and the virtues of this workshop may assist other countries who are revising their schemes of examinations in schools.

PART I

In preparation for the TES Evaluation Workshop the following operations were carried through.

(1) Selection of participants.

(2) Production of multiple choice and free response test items by participants.

(3) Editing of multiple choice test items by participants.

(4) Selection of two multiple choice pre-tests.

(5) Moderation of multiple choice pre-tests.

(6) Administration of multiple choice pre-tests for pupils.

(7) Analysis of data from the administration.

(1) Selection of participants

Three categories of participants were envisaged.

(a) Supervisory staff as personnel responsible for controlling and guiding the science programme in the field, and as potential in-service leaders for future evaluation in-service training courses.

(b) Controlling examiners, chief examiners and senior marking examiners, as personnel directly involved in the communication, execution and control of examination practices at the G.C.E. 'O'.

(c) Senior teachers who did not belong to category (b) but were recommended by the supervisors as being efficient and successful in their teaching.

The participants in each category were selected according to their subject preference, physics, chemistry and biology, and across the categories each subject group had an initial population of about 25 participants, taking into account the possibility of drop-outs before the Workshops. At the Workshop the number expected in each subject was 20.

(2) Production of multiple choice and free response test items by participants

Each participant was requested to make 40 test items and a table of specifications with a content coverage of the whole of the G.C.E. 'O' syllabus, and an objective coverage as follows:

15 of I, 10 of II, 10 of III, III', and 5 of IV.

The table of specifications portrayed no objectives in the psychomotor or in the affective domain.

With the request to make test items the participants were also sent a note on the construction of multiple choice test items, tables of specifications and simple analysis of data (see Appendix 1). They were also sent a note on the construction of structured essay and free response questions (see Appendix 2).

A time schedule of one month was given to the participants for this task, with the request that half the items be sent in after 15 days.

When the item sets were received they were identified with a code number given to each participant, and the questions typed so that the objective suggested by the writer also appeared against the question.

(3) Editing of multiple choice test items by participants

Each participant was then sent a full set of items produced by him and by his colleagues and was requested to edit these in the following manner:

(i) judge the objective category of the question and mark the choice with a circle in the set of objective categories indicated at the end of each question;

(ii) mark the correct choice with a circle for each question;

(iii) edit questions for obvious errors and ambiguities and write the corrections on the particular question.

These edited sets were returned, and formed the basis of the raw pool from which selection was made for the pre-test.

(4) Selection of the two multiple choice pre-tests

The selection was made by two supervisors per subject group. These supervisors were to be the leaders in the Evaluation Workshop and had already participated in the operations so far. These group leader supervisors then met under the guidance of the Director of the Workshop, and selected items from the raw pool, making further editing as required. They then produced two 40-item multiple choice pre-tests.

(5) Moderation of multiple choice pre-tests

These selected pre-tests were then subjected to moderation by the control examiners of the national G.C.E. 'O' examinations. Their function was mainly to ensure that the pre-test had face validity when compared with national G.C.E. 'O' examinations.

(6) Administration of multiple choice pre-tests for pupils

The sample to whom the pre-tests were administered came from the English medium and consisted of 100 final year G.C.E. 'O' pupils plus 50 first year G.C.E. 'A' pupils, for each of the pre-tests. It was necessary to include some pupils from the G.C.E. 'A' population because it was recognised that in June and July the pupils would not have covered the full syllabus. On the other hand, the G.C.E. 'A' pupils had not only covered the syllabus but had been sufficiently successful in the various subjects of the G.C.E. 'O' examination to permit them to be in the Advanced Level classes. It was recognised that only tentative extrapolation could be made with the data from the G.C.E. 'A' population.

Although translations were made of these pre-tests in the Sinhala and Tamil languages, and the tests administered to parallel populations as in the English medium, the data following from the Sinhala and Tamil media were not intended to be utilised at the Workshop, but were expected to be used with the participants as follow-up work during the course of the year, after the Workshop.

(7) Analysis of data from the administration of the multiple choice pre-test

It was recognised that the participants were mainly personnel involved in classroom evaluation activities and not those who would be taking the primary responsibility for the examination design and decision making operations at a national level.

One of the problems that these participants would meet in the analysis of data in a school situation is the smallness of number in their samples, and the differences between national indices and the indices for their particular situation. To highlight this problem, difficulty indices and discrimination coefficients were calculated both in terms of the total sample population and in terms of the individual school sample population. Further, both the school data and the total sample data were divided into the G.C.E. 'O' and G.C.E. 'A' sub-sets. The pupil performance characteristics were reflected in the standard deviations and means calculated for the G.C.E. 'O' and the G.C.E. 'A' sub-sets, and for the total sample. The analysis on a school-wise basis was done by the Evaluation Research Unit under the direction of the Director of the Workshop, and the total sample analyses were done by the supervisors who were to act as the group leaders at the Workshop. During the course of this operation, the group leaders were briefed on the implications of the analysis, and on the apparent trends and interpretations visible from the data.

The pre-workshop operations in addition to providing realistic working material for the Workshop, brought the participants to a preliminary level of awareness of the problems involved in the construction of multiple choice items, specific objectives and content; in the communication through tables of specifications; in the preliminary editing processes, including judging the categorising of items according to objectives. In addition, the group leaders had an opportunity to involve themselves in the other operations planned for the Workshop, such as the selection and rejection of test items, and the analysis and interpretation of test data.

PART II

The participants at the Workshop

There were two Directors of the Workshop, one from Ceylon and the other from the United Kingdom. There were ten other participants, in the chemistry group all of them teachers, training college lecturers or education officers in Ceylon.

The material available

1. Two sets of fixed response questions selected by the educational supervisory staff in Ceylon from approximately 400 such questions submitted during the preliminary work. Each set contained 40 pre-tested questions.

2. The data available on each question was: the content area and educational objective and the difficulty index and discrimination coefficient as determined by the pre-test.

The content area indicates the main divisions of the material of Ceylon 'O' level chemistry.

The educational objectives are divided into five categories:

1. Recall.
2. Comprehension.
3. Application (qualitative).
3'. Application (quantitative).
4. Analysis/synthesis/evaluation.

The difficulty index shows the percentage of pre-test candidates who answered the question correctly.

The discrimination coefficient shows the correlation between the performance of candidates on a particular question with that in the test as a whole (how the item separates good pupils from poor pupils).

3. Administrative requirements: two rooms, each with facilities for writing; and typing and duplicating staff sufficient to type and duplicate the product of each day's work during the evening. Computing equipment was not required.

Time available

There were 9 working days. The evaluation workshop sessions amounted to 23 hours. (The participants were engaged in a seminar which included curriculum development work in addition to the evaluation workshop.) The participants had an average of about 2 hours in each day writing questions or other evaluation work in private.

It was found essential during the operations which follow to appoint one participant as secretary.

Sequence of operations

1. Reviewing objectives. After short preliminary talks by the Directors on the aims and methods of the Workshop, participants reviewed the objectives of all the pre-tested questions. The latter served the purpose of clarifying to the participants the educational objectives in operational terms.

The first 40 questions were reviewed by general discussion under the guidance of one of the Directors. When there was divergence of opinion the view of the majority was taken. Where divergence of opinion was equal, the lower category of objective was selected, subject to the over-riding decision of the Director. The discussions were lengthy but fruitful. All participants agreed that the exercise forced them to consider the fundamental objectives of the course in terms of what they should expect their students to be able to do at the end of it.

This sort of discussion requires some skill and experience on the part of the Director. In order to give such experience to other Ceylon personnel, the Workshop was then divided into two groups each in the charge of a Ceylon participant. The two groups separately reviewed the educational objectives of the remaining 40 questions. The groups then rejoined and decisions were made on the objectives of questions where they differed from the original classification. Finally a revised specification of all 80 items was drawn up.

2. The next task was to determine:

the distribution of difficulty indices over the 80 questions;

the distribution of discrimination coefficients with difficulty indices;

the distribution of difficulty indices with objectives;

the distribution of difficulty indices with content areas;

the distribution of discrimination coefficients with content areas;

the distribution of discrimination coefficients with objectives.

To expedite the work the participants were divided into four groups. Their findings were then duplicated and given to all of them.

The main findings were:

2.1 Difficulty increased with category of objective (I/II/III/III'/IV).

2.2 The later the topic in the course, the greater the difficulty.

2.3 Calculations tended to be more difficult than other questions.

2.4 Items of medium difficulty seemed to discriminate best.

3. Selection of a 40-question test. The 80 questions were then reviewed and classified as follows:

(i) Acceptable in the original.

(ii) Acceptable after minor editing (editing not likely to affect the performance in the question).

(iii) Acceptable only after major rewriting (rewriting likely to affect performance).

(iv) Rejected.

The participants as a group went through the 80 questions several times judging the items according to the following criteria:

3.1 Duplication of items.

3.2 Too easy or too difficult (some were rejected on these grounds.

3.3 Poor discrimination - any item with a discrimination coefficient below + 0.2 was accepted only if in the subjective judgment of the participants it was suitable.

3.4 Non-functioning responses.

3.5 Ambiguity and faults in language and content (each participant acted as reviewer of a part of the questions in order to speed up this operation).

3.6 Undesirable effect on teaching.

The results of this "shredding process" show that of the 80 questions:

14 were acceptable in the original.

26 were acceptable after minor editing.

15 required major rewriting.

25 were rejected.

4. Necessary writing to meet the specification. It now remained to determine whether the 40 acceptable questions met the desired content/objective specification. It was found that some content areas were oversubscribed, objective 1 was oversubscribed and objective 3' undersubscribed. Therefore five items had to be withdrawn and five new items written in their place. The latter was done by all the participants working individually, but it was found desirable that operation 4 and subsequent operations should be undertaken by a small group of three participants plus a Director.

In this way the preferred specification was met approximately but not exactly.

5. Making a random key. It was found that there was some imbalance in the correct responses.

Response	(i)	in 10 questions.
Response	(ii)	in 9 questions.
Response	(iii)	in 13 questions.
Response	(iv)	in 8 questions.

The responses in one question, correct response (iii), were reordered to give a correct response (iv), thus giving a better distribution of responses.

6. Order of questions. During discussion it was noted that in the pre-test the questions were arranged so that the higher objectives, and more difficult questions, came at the end. It was decided to arrange the 40 selected items in such a way as to distribute the more difficult items through the test. Two or three easier items were deliberately put at the beginning and end of the test.

7. Specification of the test. The essential data on the 40 questions in the test was assembled.

Structured and free response questions

One of the objectives of the Workshop was to give instruction and practice in writing questions other than those of the fixed response type. Participants were asked to write some structured and free response questions before coming to the Workshop and guidance on how to do so was sent to them (see Appendix 2).

Not many of these questions were forthcoming, however, and most of the writing was done during the second half of the Workshop. This provided suitable work for those participants who were not being used in operations 4-7 above.

It was felt that this part of the Workshop was of at least equal importance to the first if desirable effect on teaching was to be attended through examinations.

Time was insufficient to submit the structured and free response questions to the same scrutiny given to the fixed response questions.

In conclusion

One of the most important outcomes of the Workshop was that the participants came to realise that teaching and examining were not separate operations but parts of a larger operation; and that taking part in a workshop such as this threw a new and clearer light on the objectives, methods and content of teaching, as well as of examining.

It was noted by all the participants that a workshop is not the best occasion on which to write questions. Ideas for questions probably come best when the writer is actively engaged in teaching, and much time is required to develop the idea and refine the question. This workshop was probably at its most valuable phase during the discussion, classification and assembly sessions; and at its least valuable during the question-writing sessions.

APPENDIX I

Introductory Notes for Evaluation Workshop

by J. Ratnaike

Introductory note on levels of achievement of pupils (objectives)

In the study of a particular area of content in the syllabus, pupils may be taught, and expected, to reach various levels of achievement in the content area. For example, the pupils may be able to recall thd reproduce the actual teaching given, and no more. This, of course, will not tell us whether the pupil has understood what he is repeating.

On the other hand, a higher level of achievement would be if he is able to apply qualitatively or quantitatively what has been taught, in a situation somewhat different to what he has met before in the classroom. As teachers, we would then recognise that the pupil has shown that he has "understood" what has been taught. We would call this level "higher" in terms of the purposes for which we are teaching, and in the sense that it involves more complex operations on the part of the pupil than merely recalling.

Pupils may even have achieved a still higher level, of being able to recognise problems, select data, draw general conclusions from data, test hypotheses - in other words use the methods of the scientists in his work.

These are levels that we as teachers perceive in our day-to-day work with pupils.

These levels of achievement or objectives, as they are called, have been expressed in various ways. Perhaps the most comprehensive classification available today is The Taxonomy of Educational Objectives, Benjamin S. Bloom et al., Longmans Green and Co. (1956). Given below is a brief summary of this classification. To find out whether the pupil has achieved one or other of these levels, we provide him with an opportunity to respond at the various levels - in other words we ask him questions. The questions included as examples below attempt to find out at which level of achievement the pupil can respond.

SUMMARY OF BLOOM'S CLASSIFICATION OF THE COGNITIVE DOMAIN

LEVEL 1. The pupil can provide responses involving primarily RECALL of such items as specific facts, terminology, conventions, trends, classifications, criteria, methodology, principles, generalisations.

EXAMPLES OF QUESTIONS REQUIRING PUPIL RESPONSES AT LEVEL 1

Content area: Air/combustion.

The amount, by volume, of carbon dioxide in the atmosphere is about:

(i) 30 parts per 100
(ii) 30 parts per 1,000
(iii) 3 parts per 1,000
(iv) 3 parts per 10,000

Content area: Nitrogen peroxide/equilibrium.

Which one of the following crystalline substances, when heated strongly in a hard glas test tube, would NOT give brown fumes?

(i) Potassium nitrate
(ii) Lead nitrate
(iii) Barium nitrate
(iv) Copper nitrate

LEVEL 2. The pupil can provide responses involving more than simple recall. An element of COMPREHENSION is required, and may be of the types such as the following: translation of mathematical verbal material into symbolic statements and vice versa; interpretation of data and statements (including simple explanations, summaries, comparisons); extra- and intrapolation from data, statements (i.e. extension of trends beyond the given data); determination of effects; corollaries.

EXAMPLES OF QUESTIONS REQUIRING PUPIL RESPONSES AT LEVEL 2

Content area: Air/combustion.

It is possible to recognise some similarities between the burning of common home fuels, and respiration. However, it is NOT true to say that, in the cases of both burning and respiration:

(i) the gaseous products are similar in properties.
(ii) the substances burned, and foodstuffs, have at east two chemical elements in common.
(iii) oxygen is required for the reactions to take place.
(iv) most of the energy from the change appears as chemical potential energy of the products.

Content area: Nitrogen peroxide/equilibrium

Nitrogen peroxide containing 21% nitrogen dioxide is passed, at one atmosphere pressure, through a long non-porous inert glass tube completely surrounded by an apparatus A. The gas emerging at the end of the glass tube was found to have a composition of 12% nitrogen dioxide. Which one of the following may be said of the apparatus A?

(i) A is a heater.
(ii) A is a cooler.
(iii) A is a filter for nitrogen dioxide.
(iv) There is insufficient data to identify A.

LEVEL 3. The pupil can provide responses involving the APPLICATION of abstractions in particular concrete situations. The abstractions may be in the form of general ideas, rules of procedure, general methods, technical principles, theories. The applications may include numerical applications as well. (For convenience we may indicate the level of numerical application with 3*.)

EXAMPLES OF QUESTIONS REQUIRING PUPIL RESPONSES AT LEVEL 3

Content area: Air/combustion.

Which pair of the observations given below, made when magnesium was ignited in air, offers the best evidence to support the contention that a chemical change has taken place?

(i) Heat evolved; colour change.
(ii) Colour change; light evolved.
(iii) Light evolved; volume change.
(iv) Volume change; mass change.

Content area: Nitrogen peroxide/equilibrium.

When a mixture of chlorine and helium (called gas mixture X) is in a container at room temperature and 1 atmosphere pressure is passed through a series of long inert porous tubes, it is observed that the emerging gas has a relative density different to that of the original gas. Which one of the following (and for what reason) would you expect when instead of gas X, nitrogen peroxide is passed through the above set-up, under the same conditions of temperature and pressure?

(i) No change in the relative density because gas mixture X is a mixture of lighter and heavier components whereas nitrogen peroxide is a substance.

(ii) No change in the relative density because gas mixture X is a mixture of lighter and heavier components, whereas the components in nitrogen peroxide are in dynamic equilibrium.

(iii) A change in the relative density because, like gas mixture A, nitrogen peroxide contains lighter and heavier molecules.

(iv) A change in the relative density, because the heavier molecules of dinitrogen tetroxide decompose into the lighter molecules of nitrogen dioxide as they are separated.

EXAMPLES OF QUESTIONS REQUIRING PUPIL RESPONSES AT LEVEL 3[*]

Content area: Air/combustion.

What volume of oxygen, at S.T.P., would be expected to be required for the complete combustion of 240 g (one gramme molecular mass) of a fuel oil of molecular formula $C_{17}H_{36}$?

(i) 26 x 22.4 litres
(ii) 26 x 16 x 2 litres
(iii) 35 x 22.4 litres
(iv) 35 x 16 x 2 litres

Content area: Nitrogen peroxide/equilibrium.

Two gramme molecular mass of copper nitrate is to be heated in a hard glass vessel. Any gas produced will be collected over water at 20°C. Assuming that all the S.V.P. and S.T.P. corrections may be neglected, what volume of gas would be expected to be collected? The equation describing the change may be taken as:

$$2Cu(NO_3)_2 \longrightarrow 2CuO + 4NO_2 + O_2$$

(i) 5 x 22.4 litres
(ii) 4 x 22.4 litres
(iii) 2.5 x 22.4 litres
(iv) 1 x 22.4 litres

LEVEL 4. The pupil can provide responses involving ANALYSIS, SYNTHESIS and EVALUATION. This would include such operations as: recognition of explicit and implicit relations, structure; recognition of unstated assumptions; discriminations of facts from hypotheses; organisation of ideas, statements, data; propositions of methods for testing hypotheses, theories; design of experiments; recognition of logical accuracy, consistency; comparison of theories, generalisations.

(Note: Bloom's Taxonomy actually subdivides Level 4 into three levels, 4, 5, 6 for each of the components.)

EXAMPLES OF QUESTIONS REQUIRING PUPIL RESPONSES AT LEVEL 4

Content area: Air/combustion.

Two theories were suggested to explain what happens when substances burn in air. Neither assumed negative weight.

Theory L: When the substance is burned, it combines with a part of the air.

Theory S: When the substance is burned, the air takes up something that leaves the substance, until the air is saturated with it.

JUDGE Whether each of the observations given below:

(i) is more direct evidence for Theory L than for Theory S.

(ii) is more direct evidence for Theory S than for Theory L.

(iii) supports both Theory L and Theory S.

(iv) supports neither Theory L nor Theory S.

Observations

(a) A coconut oil lamp put into a large bottle, and the lid closed, goes out before the oil is consumed.

(b) The white product formed when magnesium burns in air weighs more than the original magnesium.

(c) When the red product, formed on heating mercury in air, is reheated in the used-up air, the original properties of the air are restored.

(d) Much heat and light are given out when magnesium burns.

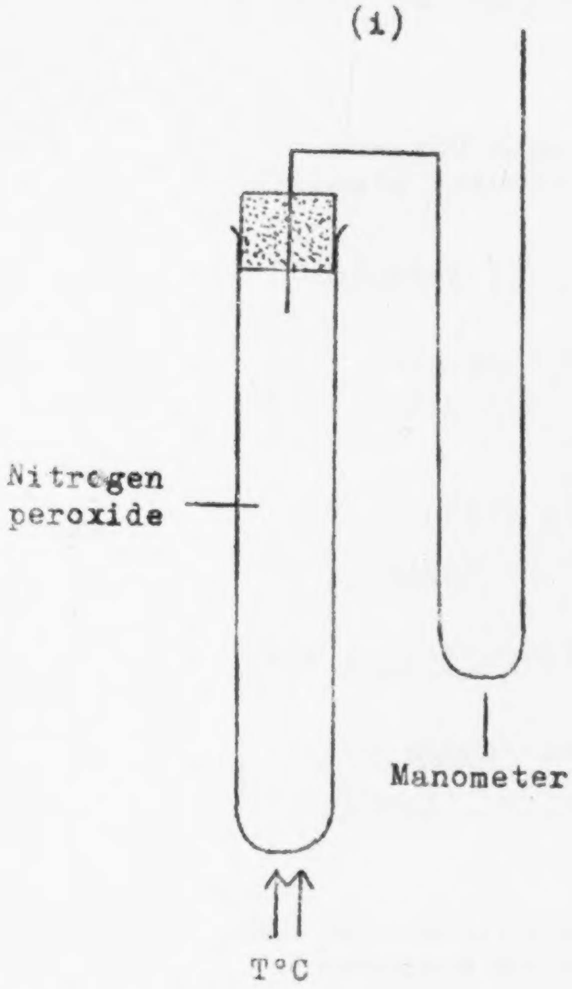
(e) Turpentine burns with a more reddish flame in chlorine than in air.

Content area: Nitrogen peroxide/equilibrium.

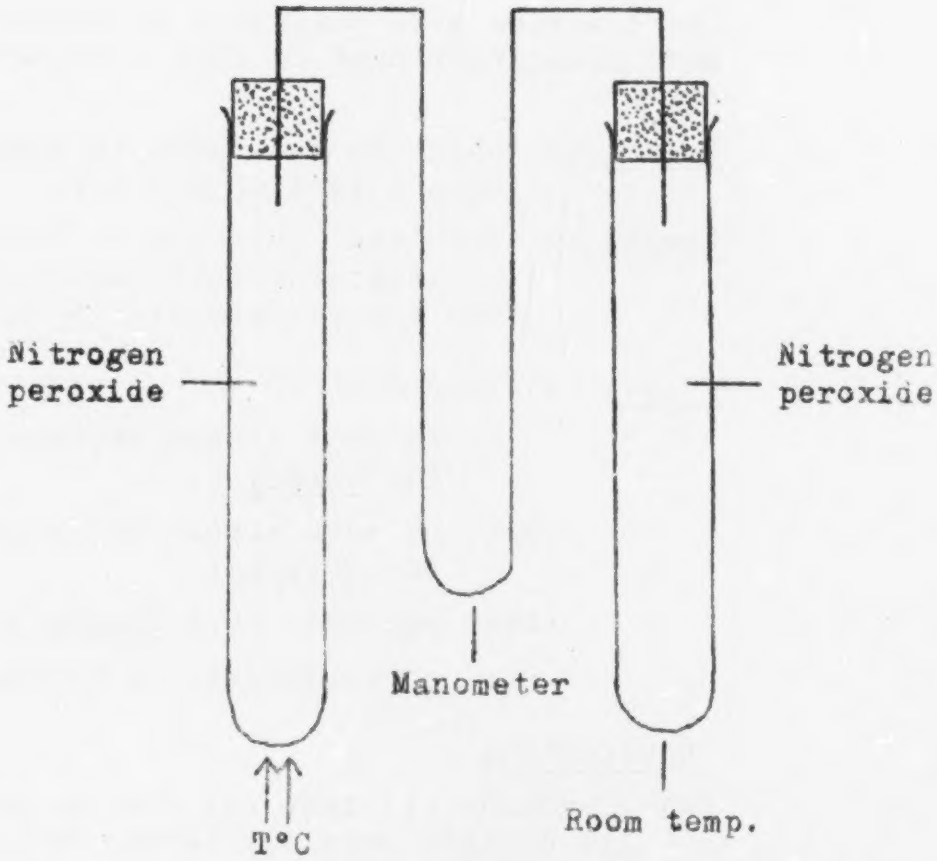
To study the effects of dissociation in nitrogen peroxide using changes in pressure when the gas is heated through the range 40°C to 80°C, which one of the following may be considered the most suitable?

(In the diagram ↑↑ indicates heated in a constant temperature bath. Each of the gas containers has a volume 100 ml.)

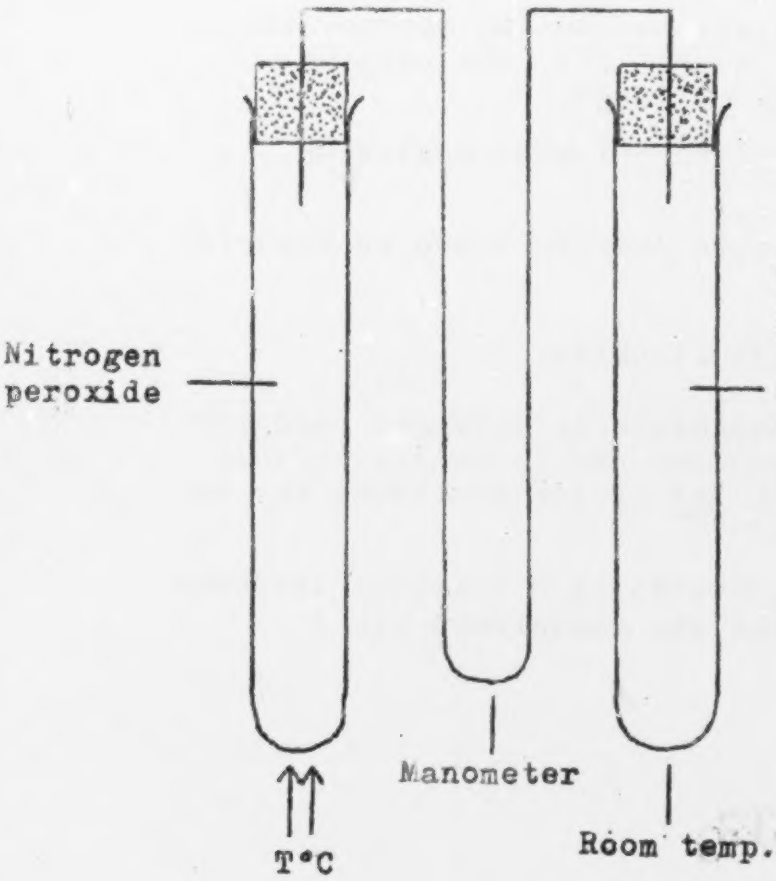
(i)


Nitrogen peroxide
Manometer
T°C

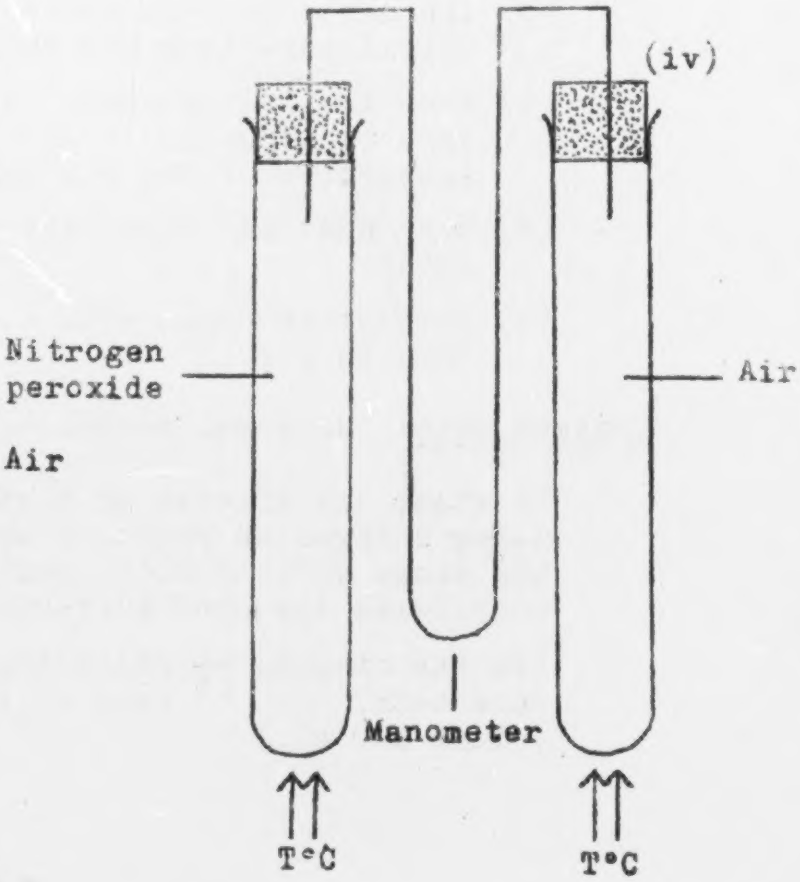
(ii)


Nitrogen peroxide
Nitrogen peroxide
Manometer
T°C
Room temp.

(iii)


Nitrogen peroxide
Air
Manometer
T°C
Room temp.

(iv)


Nitrogen peroxide
Air
Manometer
T°C
T°C

Some comments on the design of the questions illustrating the levels of achievement

LEVEL 1

Content area: Air/combustion.

Only the order of magnitude is required, particularly as there are significant variations from place to place. The fact has been given to the pupils in the classroom. A possible (arithmetic) difficulty may arise if the pupil has been given the value of the carbon dioxide concentration as a % (0.03%). Is there any use in asking such a question?

Content area: Nitrogen peroxide/ equilibrium.

Again pupils have met these facts in the classroom. The question while being simple recall, still covers somewhat more content (the action of heat on nitrates) than the previous example. Pupils may apply the empirical generalisation pertaining to the heavy metal nitrates - in which case the question may be classified as Level 2. The rule often used is to classify in the lower (in complexity) category, if there is an ambiguity.

The question may be designed to cover even more nitrates by using combinations of nitrates, and wording the question, "Which of the following pairs, when heated separately ....?" In situations where the substances involved react independently even in the mixture form, the question may be so worded.

Note that as the question is in the negative, the NOT is emphasised.

LEVEL 2

Content area: Air/combustion.

Since the question requires a comparison to be made between the burning of common fuels, and respiration, using factual knowledge already given to the pupils in the classroom, it is classified in Level 2. The word "chemical" has been purposely introduced in choice (ii) because of the possible ambiguity of the word "element" if used alone. "Elements in common" may in ordinary language also mean properties or characteristics in common. The critical difference of the manner of transformation of energy in the two cases is being tested for. For homogeneity it is not necessary for all the choices to be of equal length. In this example choices (ii) and (iv) are long while (i) and (iii) are short.

Content area: Nitrogen peroxide/equilibrium.

The question involves simple interpretation of data. It is therefore classified in Level 2. The pupil has to comprehend that if the % nitrogen peroxide has reduced in the system, it is due to a reduction in temperature. The question may be argued to be of Level 3. But actually the question is merely a reversal of what has been taught in class - that when the temperature is reduced the nitrogen dioxide reduces - and stated in reversal it would be, if the % nitrogen dioxide has reduced it is due to fall in temperature. So it could not be in Level 3, which involves application of a higher order of sophistication. (The above reversal is true if no other factors affect.)

LEVEL 3

Content area: Air/combustion.

The general idea of the criteria used to recognise chemical change is being applied in the situation of magnesium burning in air. Hence the question is classified in Level 3. The question requires that the criterion used is the one that is most "sufficient" (in the sense of "necessary" and "sufficient"). It does not demand that the selection should be uniquely and absolutely "sufficient". There may be some doubt as to whether the question should be classified in Level 4, since the operation of evaluation is taking place, and a judgment is being made. Since the process of judgment in this case is not sophisticated, and applying the rule stated earlier regarding ambiguities in classification, the lower level classification has been used. It should be noted that in answering any multiple choice question, or for that matter any question, the pupil has to make some kind of judgment in the process of answering.

The technique of coupling done in the choices is a way of bringing into the four-choice item more than four characteristics. It also uses the idea that a characteristic coupled to a second may make the best evidence available while the same characteristic coupled to some other would not do so.

Content area: Nitrogen peroxide/equilibrium.

This is a more difficult application question, since it applies the ideas of diffusion, as well as those of dynamic equilibrium. Further, a decision has to be made (a) as to whether there is a density change and (b) for what reason. Sometimes these two may be separated to form two different test items, the first to test for (a) and the second for (b) to find out whether the choice of (a) has been made for the correct reasons.

The choice (iii) is a little shorter than the choice (iv) with which it is coupled, and is therefore visually inhomogeneous.

Choice (iii) might have been worded, "A change in density because, like gas X, nitrogen peroxide contains lighter (nitrogen dioxide) and heavier (dinitrogen tetroxide) molecules", but this could cause ambiguity because of the possible misinterpretation of the statement to mean that gas mixture X contains nitrogen dioxide and dinitrogen tetroxide.

Note that in choices (ii) and (iii) essentially the same reason is being given, first to justify no density change and then to justify density change. But by changing the wording in the two cases, the reasons have been made plausible.

LEVEL 3*

Content area: Air/combustion.

This question requires the numberical application of the 22.4 litres idea. Its classification in Level 3* is obvious. There are many ways of arriving at the answer. One of the shortest is $C_{17} \longrightarrow 17CO_2 \longrightarrow$ 17 moles oxygen and $H_{36} \longrightarrow 18H_2O \longrightarrow 18 \times \frac{1}{2}$ moles oxygen. Hence (17 + 9) = 26 moles of oxygen, hence 26 x 22.4 litres.

The choices include common errors that may be made. Choice (ii) has 35 = (17 + 18), where the realisation that each molecule of H_2O would take in $\frac{1}{2}$ mole oxygen has been missed. The other two choices involve the common errors of using the molecular mass of 16 (for oxygen) and getting the answer as a mass rather than as a volume. Other possible choices may be made up by presenting the common error or 36 atoms of hydrogen giving only 18 molecules of water and not 36, thereby generating two "volume" choices ($36 \times \frac{1}{2} + 17$) moles oxygen and (36 + 17) mole oxygen, as well as more "mass" choices. For a particuler examined population, judgment will have to be made as to which sequence of choices would be more discriminating. To have homogeneity, the "26" choices and the "35" choices have been placed in that order.

Content area: Nitrogen peroxide/equilibrium

The classification of this question in Level 3* is obvious too. But the question not only requires the application of the idea of the 22.4 litres, but also the fact that nitrogen peroxide is soluble in water. Thereby the only gas collected is oxygen. Choice (i) is from the sum of the moles of oxygen and nitrogen peroxide. Choice (ii) considers only nitrogen peroxide. Choice (iii) would be the answer if only one mole of the nitrate were considered, as a result of the association of 1 mole of gas having a volume 22.4 litres. The assumptions are given not only to simplify the arithmetic, but also to avoid the complications of the larger and unspecified amount of dinitrogen tetroxide present in the nitrogen peroxide at S.T.P. as compared with that at 20°C, and the S.V.P. arising from the dissolved nitrogen peroxide. The two gramme molecular mass is also given to simplify the arithmetic. The choices have been written in decreasing order of magnitude instead of randomly.

LEVEL 4

Content area: Air/combustion

The question involves judging evidence in support of two theories (in fact the oxygen and the phlogiston theories). Hence the question is classified in Level 4. It should be noted that since Theory L is more comprehensive than Theory S, one would not find evidence which would support Theory S, but not Theory L. So really this is a three-choice question - but pupils who have reduced this to a three-choice question have analysed the problem in the abstract. Note that some of the observations are similar to the ones commonly given in some textbooks as unique evidence for Theory L, but which are in fact not so.

Content area: Nitrogen peroxide/equilibrium.

Analysis of the experiment and the design considerations involved are being tested here, so that it is classified in Level 4. The pupil has to analyse the situation and conclude that there could be "Charles' Law of facts" as well as "dissociation effects" that could provide for changes in pressure in the experiment. The answer involves one that will neutralise the "Charles' Law effect". Many other techniques may, of course, be used in this investigation but the point being tested here is the examinee's ability to decide which of the choices given would provide the best design for the investigation.

In which level of achievement (objective) would you place the following questions?

(1) A sky-rocket firework gave a bright crimson light. A salt of which one of the following substances may have been used to produce this?

(i) Strontium (ii) Potassium (iii) Barium (iv) Sodium

(2) A substance effective in removing iron rust stains on cloth is:

(i) Alcohol
(ii) Kerosene
(iii) Sodium chloride solution
(iv) Oxalic acid solution

(3) Chlorine may be prepared by using which set of the following chemicals?

(i) Granulated zinc and concentrated hydrochloric acid.

(ii) Potassium permanganate and concentrated hydrochloric acid.

(iii) Concentrated hydrochloric acid and concentrated sulphuric acid.

(iv) Sodium chloride and concentrated sulphuric acid.

(4) Which one of the following may be taken as evidence for judging whether the given material is NOT a pure substance?

(i) When heated in a test tube, the portion near the flame always melted first.

(ii) Under the microscope the substance was observed to be made up of particles of different sizes.

(iii) Larger particles of the substance were found to have a higher density than smaller particles.

(iv) 10 g of the substance all dissolved in 50 g of water at 30°C.

(5) In carbon monoxide 12 g of carbon combined with 16 g of oxygen, while in carbon dioxide 12 g of carbon combined with 32 g of oxygen. Is this statement:

(i) an observed fact supporting Dalton's Theory?

(ii) an assumption made in Dalton's Theory?

(iii) an observed fact not involved in Dalton's Theory?

(iv) an assumption not involved in Dalton's Theory?

(6)

Pressure (atoms)	10	100	300	1000	Temp. 200°C
% conversion	50	82	91	99	

To which of the following reactions could the behaviour shown in the above table NOT apply?

(i) $CO + 2H_2$ CH_3OH

(ii) $N_2 + 3H_3$ $2NH_3$

(iii) $N_2 + O_2$ $2NO$

(iv) $3C_2H$ C_6H_6

(7) Which one of the following solutions, with the pH values given, would you expect to be most sour?

(i) pH 1 (ii) pH 3 (iii) pH 7 (iv) pH 9

(8) It is thought that atoms combine with each other so that the outermost shell acquires a stable configuration of 8 electrons. If stability were attained with 6 electrons rather than with 8, what would the formula of the stable fluoride ion be?

(i) F^- (ii) F^+ (iii) F^{3+} (iv) F_2

(9) The percentage of dissolved matter in sea water is about:

(i) 0.03% (ii) 0.3% (iii) 3% (iv) 30%

(10) When 10 ml hydrochloric acid is added to 10 ml silver nitrate, what will the resulting mixture contain?

(i) Excess hydrochloric acid.
(ii) Excess silver nitrate.
(iii) No excess reagent.
(iv) Impossible to say.

(11) A white crystalline material which does not react chemically with water, when extracted with excess water, left a residue A. The water extract left a residue B when evaporated to dryness. The residue A was further extracted with excess water, and this extract left no residue on evaporation.

Which one of the following conclusions may be drawn from the above observations only?

(i) The material is a pure element.
(ii) The material is a pure compound.
(iii) The material is a mixture.
(iv) One cannot say any of the above with certainty.

(12) Bright pieces of metals indicated were placed in the water solutions of the compounds indicated. All the pieces of metals appeared to behave in a similar fashion except one. Which one was this?

(i) Iron in a solution of copper sulphate.
(ii) Copper in a solution of magnesium sulphate.
(iii) Zinc in a solution of copper sulphate.
(iv) Iron in a solution of silver nitrate.

(13) Compared with another atom of atomic mass 12 and atomic number 6, the atom of atomic mass 13 and atomic number 6:

(i) contains more neutrons.
(ii) contains more protons.
(iii) contains more electrons.
(iv) is a different element.

(14) When brass pins are tin plated using stannous chloride solution, the element that is reduced is:

(i) Tin (ii) Copper (iii) Zinc (iv) Brass

(15) A batch of pupils going in a bus had to keep the windows closed for several hours because of rain. Samples of air inside the bus were found to have a high concentration of carbon dioxide, which was probably:

- (i) gases escaping from the exhaust pipe of the bus.
- (ii) the oxidation of carbohydrates, proteins, fats.
- (iii) the atmosphere outside.
- (iv) the burning of the petrol in the bus.

Writing Test Items

Writing test items is essentially a creative art. Broad principles and techniques may be suggested, but these will not guarantee the production of good test items. It is the item-writer's judgment and creative ability to make use of, or even ignore, these principles and techniques, that will determine whether an item will be good, mediocre or poor.

However, no amount of knowledge of techniques will make up for a lack in subject matter. Also a knowledge of techniques and of subject matter alone will not make up for a lack of knowledge of the characteristics of the examinee population.

The comments below refer essentially to test items of the multiple choice form, which is one of the commonest of the various objective test forms. This does not by any means imply that the objective test is the best of the testing devices available. The following comparison will make this clear.

Comparison of essay-type (free response-type) and objective-type tests

Preparation of test item:	It is commonly thought that essay-type questions are easy to construct, whereas objective questions are relatively more difficult. However, to construct good, searching, essay-type questions, structured for reasonable reliability in marking, of levels of achievement 3 and 4, could be as difficult as constructing objective tests of the same levels. Constructing factual recall essay-type questions is easier than constructing factual recall objective-type questions.
Sampling of content:	By its nature, sampling in an essay-type test is often limited. In objective-type tests, extensive sampling can be obtained fairly easily. The sampling in essay-type questions may be extended by framing questions that bridge different areas of

content - say of the "compare and contrast" kind - but these are usually rather difficult to construct, especially at levels 3 and 4.

Measurement of levels of achievement: Both types are capable of measuring all the levels of achievement in the cognitive domain.

Preparation and mode of study by pupils: There is some experimental evidence to suggest that when pupils prepare for essay-type tests they concentrate on larger units of subject matter and utilise study methods that favour working out trends, relationships, interpretations and organisations. In preparation for essay-type tests, pupils tend to make summaries, and practice communicating verbally in an organised way. For objective-type tests, pupils tend to produce study habits that concentrate on details and specifics, rather than on the larger aspects of the subject.

Nature of response: Pupils organise their responses in essay-type questions. In effect, these questions are multiple choice questions of a very large number of choices, none of which are given. In the objective-type test, the responses are given. The pupil does not have to design or originate the answer. He has only to recognise and select the answer. Lack of fluency of expression is not a factor in objective-type tests, although a good command of the language is required to discriminate among the choices. Original thinking is very difficult to be tested with objective-type tests, whereas in essay-type tests it is quite easy.

Time for response by pupils: The time of answering in essay-type tests is of course much longer than (for comparable content coverage) in objective-type tests. However, reading time in the objective-type tests is considerable.

Guessing by pupils: In essay-type tests, successful guessing is a very minor problem, and is often the fault of the examiner for not using a carefully designed table of specifications for the test, and thereby repeating questions covering the same content area. On the other hand, guessing could be a major problem in the objective-type tests, especially in tests where time of answering is barely enough.

Scoring of pupil responses:	In essay-type questions, scoring is difficult, time-consuming and somewhat unreliable, even with a carefully designed marking scheme, unless elaborate structuring has been built into the design of the essay-type questions. In the objective-type tests, scoring is simple, rapid and highly reliable.

This comparison reveals the fallacy of the argument that one type of test item is unquestionably superior or inferior to the other. Each has unique advantages, and unique limitations. There are testing needs for which each is particularly well suited. The experience of several nation-wide examination research organisations is that examinations should contain both types of question, in a proportion determined by the needs of the particular test situation.

The following few rules of construction are not comprehensive, but are sufficient to start producing questions. Further sophistication and refinement will come with experience.

Rule (1) Use either a direct question or an incomplete statement as the item stem.

EXAMPLE: Direct question

When you blow on a candle, it is put out. Which one of the following would be the best reason to explain this?

(i) Carbon dioxide in the exhaled air is a non supporter of combustion.

(ii) Moisture in the exhaled air moistened the candle wick.

(iii) Candle wax vapour cooled below its ignition point.

(iv) Blowing caused a change in the atmospheric pressure.

Incomplete statement

Reaction between neutral barium chloride and sodium carbonate goes to completion because:

(i) a gas is formed.

(ii) the reaction is reversible.

(iii) barium carbonate is insoluble.

(iv) sodium chloride is more stable than sodium carbonate.

Although both kinds of stems are suitable, beginners in test construction seem to produce fewer technically weak items when they try to use direct questions than when they use the incomplete statement form. Because of the specificity of the direct question kind, it induces the item writer to produce more specific and homogeneous responses. When an incomplete statement is used, the writer's point of view may shift as he writes successive responses. This can confuse the examinee about the real point of the item. It is also usually easier for the writer to express complex ideas, especially those requiring qualification, as a direct question. The need to have a completion at the end of an incomplete statement restricts the writer somewhat. He is not free to arrange phrases or words to produce the clearest possible statement. Further, the writer of a direct question usually states more explicitly the basis on which the correct statement is to be chosen. However, in the direct question kind, the number of words are usually more than in the incomplete statement kind, and this is a factor to be considered in the context of reading time, which, in any case, for a multiple choice test, is much more than for a similar (content and objective coverage) essay-type.

Rule (2) Generally include in the stem any words that must otherwise be repeated in each response.

EXAMPLE: Under incomplete statement above, the word "because". However, it is not always possible or desirable to eliminate all words common to the responses. These words have to be left in the responses to make grammatical sense, or for clarity. (See example under levels of achievement, Level 2.)

EXAMPLE: When potassium chlorate is heated with manganese dioxide, a reaction takes place. Manganese dioxide, in this reaction, may best be described:

(i) as an inert component, because its mass does not change during the reaction.

(ii) as an oxidising agent, because oxygen is liberated in the reaction.

(iii) as a reducing agent, because the oxygen-free potassium chloride is produced from potassium chlorate.

(iv) as an accelerator, because oxygen is given out at a lower temperature than with potassium chlorate alone.

(See also example under Rule (3).

Rule (3) To begin with, avoid too many negatively stated items.

For one, such questions are more difficult to make because three of the responses have to be correct responses. For another, examinees may be confused. Having been accustomed to selecting a correct response, they may find it difficult to remember to select an incorrect response. If a negatively stated stem is used, the negative should be clearly emphasised, say, with capitals, thick letters, or italics.

EXAMPLE: In a balanced chemical equation representing a definite chemical reaction, we would expect to find certain factors on one side the same as on the other. Which one of the following would NOT be the same on the two sides?

(i) Total number of molecules on the right.
(ii) Total number of atoms on the right.
(iii) Sum of the atomic masses on the right.
(iv) Sum of the atomic numbers on the right.

(Note: The repetition of "on the right" is essential for clarity - see Rule (2).)

Rule (4) Avoid making the correct response obvious due to stereotyped phrasing or grammatical clues or due to visual inhomogeneity, such as making it very precise and long (by qualifying it to make it accurate)

EXAMPLE: Sliced oranges are an excellent source of:

(i) Protein.
(ii) Starch.
(iii) Vitamin A.
(iv) Vitamin C, if freshly sliced.

Rule (5) Use "none of these" or "none of the above" only in items to which an unambiguously correct answer may be given.

It should not be used as a "space filler" because a fourth choice cannot be produced. Used with discrimination, it can be a very good distractor.

EXAMPLE: Hydrogen was passed over the heated (about 600°C) solid products shown in items (A) and (B) below. Which one of the following probably took place?

(i) $XO + H_2 \longrightarrow X + H_2O$
(ii) $X_2O + H_2 \longrightarrow 2X + H_2O$
(iii) $2X + H_2 \longrightarrow 2XH$
(iv) None of these.

(A) Residue after heating magnesium carbonate.

(B) Residue after heating copper (II) carbonate.

(See also items in the sequence for practice in classifying levels of achievement).

Rule (6) If an item deals with the definition of a term, it is often preferable to include the term in the stem and present alternative definitions in the choices.

Thereby it provides more opportunities for attractive plausible choices, and tends to reduce the chance of choosing the correct response by verbal association. (In any case, stereotyped phraseology should not be used in the stem or in the choices.)

EXAMPLE: Compare (a) and (b) below.

(a) Which of the following names would best fit the group of complex organic compounds that are found in small quantities in natural foods, and are essential to normal nutrition?

(i) Nutrients.
(ii) Calories.
(iii) Vitamins.
(iv) Minerals.

(b) Vitamins may be best described as:

(i) complex substances necessary for normal human development, and which are found in small quantities in certain foods.

(ii) complex substances prepared in biological laboratories to improve the nutrient qualities of ordinary foods.

(iii) substances extracted from ordinary foods, which are useful in destroying germs in the body.

(iv) highly concentrated forms of food energy which would be used only on a doctor's prescription.

The Table of Specifications

In our usual school tests we generally wish to test for the level of achievement in particular content areas. This means that we have at least two dimensions to our tests - the level dimension and the content area dimension.

To ensure that we have sampled the content areas (content validity) and also the levels of achievement (objective validity) effectively, it is often helpful to draw up a table of specifications for the examination we set, so that any paucity in either dimension is immediately seen by us.

An example of such a table of specifications is given in Part II - Supporting Evidence, 1. This design has been used on a national scale for the last four years. The table should be on one sheet of paper, in which case only one content column is required.

The examination for which the table of specifications has been drawn up has both types of questions - objective (type I) and essay (type II). There are 40 items in the objective-type test and these are numbered from 1 to 40 at the top of the table. There are 10 items of the essay-type, and these are numbered from 1 to 10, and follow the objective question numbering at the top of the table.

Vertically down the left, the various content areas are indicated. The syllabus has been broken up into 27 content areas. This is an arbitrary subdivision. Any convenient number based on a realistic teaching unit breakdown may be used.

Horizontally at the bottom of the table are the levels of achievement, or objectives, 1, 2, 3, 3', 4. These same levels of achievement, or objectives, are indicated vertically down the right of the table.

At a glance, we can see what this content and objective coverage are and which areas of content have not been covered or have been covered more than once. In this particular examination paper, an attempt has been made to cover almost completely the whole of the content area in the factual recall objective questions, so that the X's go diagonally down the table up to item 14, on sheet A.

After items in each of the areas at various levels of achievement have been prepared, we have to decide what kind of weightage we will give in a particular test to particular content areas and objectives. These would be the specifications of the examination paper. We can then draw up a rough table of specifications and place our items accordingly. Usually, before the examination paper is finalised, modifications in the weightage and/or the content areas that are to be tested at which particular level of achievement are required. With the finalising of the examination paper we will also have the finalised examination table of specifications which will describe the content and level of achievement coverage of the examination.

If we superimpose two or more such tables from different tests we can see, among other characteristics, whether we have repeatedly sampled the same content areas in the items at levels 2, 3, 3', 4 (assuming that in each test we had designed to cover the whole content area for items at level 1).

Guessing in objective tests

The possibility of guessing mars all objective tests. One of the earliest corrections for guessing is as follows:

$$C - \frac{W}{n-1} = \text{Votaw's Correction for Chance Success}$$

C = number correct
W = number wrong
n = number of choices in item

In a four-choice item one would subtract one third of the number incorrect from the number correct. This formula assumes that those who do not know the answer select the correct answer only by blind guessing. But if the pupils select the answer as a result of incorrect information, this formula would over-correct. It is the experience of examiners that pupils very rarely use such blind guessing if sufficient time is given for the test. Very often pupils eliminate one or two choices from knowledge, and may guess in their decision on two choices. This is supported by the fact that only very rarely are clearly incorrect choices found to be attractive to pupils.

The usual procedures to reduce blind guessing are: (a) to instruct pupils against guessing, (b) to construct likely or plausible incorrect choices based on common errors and partial information, and (c) to allow sufficient time.

As far as the first procedure is concerned, its effects are difficult to estimate. Some will guess with or without such directions. Others will be made unnecessarily cautious. Personality traits enter into the situation and the aggressive pupil will steal a march on the submissive.

Making wrong choices plausible is a necessary feature of constructing good items. It seems that, to the extent plausibility in the incorrect choices is obtained, by making them only slightly incorrect, the effectiveness but not necessarily the amount of blind guessing tends to be reduced. If the inadequate pupil finds all the choices equally plausible, he will be guessing blindly, and chance guessing will increase.

Sufficient time is vital to any test that does not presume to be a speed test. Very often the "last moment rush" is the time when much blind guessing takes place. If the time is insufficient, chance guessing will increase.

Several testing services do not apply a guessing correction at all, while others, such as the Educational Testing Service, instruct pupils to guess intelligently when not sure, and use a guessing correction. Their instructions state:

> "You may answer questions even when you are not perfectly sure that your answers are correct, but you should avoid wild guessing, since wrong answers will result in a subtraction from the number of your correct answers."

For the usual class tests, when sufficient time is given, evidence seems to suggest that the rank order of the pupils is not significantly altered whether a guessing correction is applied or not.

The use of test data for the improvement of tests

Among the many purposes for which we use tests are the following: to discover the level of achievement of pupils in particular content areas; to discriminate between good and poor pupils.

This means that we should try to determine whether, and how well, our test items are capable of performing these functions, so that they may be improved upon for future use.

If some of the incorrect choices in multiple choice test items are unattractive to many pupils they are not serving their function. A simple way to determine this is to find out how many (as a percentage) have chosen each of the choices in the test item. If the frequency is very small for a particular choice, it would be a non-functional choice and would probably need to be removed or improved upon. A large number choosing a particular incorrect choice may indicate high discrimination or it may be that ambiguity in the question is causing pupils to choose it. In either case it is worth studying the item to find out the reason. It should be noted that the results are true for the particular examinee population we have used the items on.

Tests we give usually have some questions that are easy (many pupils get the correct answer or score high marks), and some that are difficult. A simple way by which we can determine whether a particular test item has been difficult or easy, for a particular examinee population, is by using the following index:

$$\text{Index of difficulty} = \frac{\text{Number of pupils who get item correct}}{\text{Total number attempting the item}}$$

The index may be expressed as a fraction or decimal, or as a percentage. A high value would indicate that the question was easy for the particular examinee population, and a low value, that the question was difficult.

Very often, especially with objective tests, all questions are compulsory, so that the denominator in the above fraction would be the total number in the examinee group. (Using this index for essay-type questions is a little more complicated.)

We could have a similar index to tell us whether a test item can separate out the good students from the poor. For this we separate out on the basis of the total marks of the whole test the top 50% from the bottom 50% of the pupils (that is, we rank the pupils in order of their performance from the first to the last position, and separate out the top 50% and the bottom 50%). Then we find out how many in the top group got a particular test item correct, and how many in the bottom group got the same item correct. (A better estimate may be obtained if the top 33% and the bottom 33% are compared.)

Then the index of how the test item separates out the good pupils from the poor (discrimination index) is given by:

Number who got the item correct in the top group

MINUS Number who got this item correct in the bottom group

DIVIDED BY Number in the group

The number in a group will be, according to the above separation, 50% of the pupils being examined.

This index may also be in the form of a fraction or decimal, or as a percentage.

If everybody in the top group got the particular item correct, and nobody in the bottom group got the item correct, then the index would be:

$$\frac{n_T - 0}{n_T} = 1 \text{ or } 100\%$$

This would be an ideally discriminating question. Usually not all the members in the top group get the item correct, and some in the bottom group do get the item correct. So the index will be a fraction or less than 100%. The closer the value is to 1 or 100% the more discriminating. One way of making the items more discriminating is to make the wrong choices as close to the correct choice as possible (or as plausible as possible) so that only those with a clear understanding will be able to select the correct choice.

Sometimes one may find that the discrimination index turns out to be negative. This means that more pupils in the bottom group than in the top group have got the item correct. This usually means grave ambiguity in the item, in which the better pupils perhaps see more in the stem or in one of the alternatives, whereas the poorer pupils, because of their lack of ability or knowledge, are unaware of the finer interpretation that the better pupils see. Of course, one could have discrimination indices equal to zero, when about equal numbers in the top and bottom groups get the item correct. It would not be worthwhile having items of zero discrimination in a test, and of course it would be tragic to have questions of negative discrimination.

There is evidence that the questions of medium difficulty (between about 40% to 60%) are the ones that can discriminate best. The very difficult and the very easy questions do not seem to be able to discriminate very well. This is probably because the very difficult items do not engage a sufficient number of pupils. In the very easy items, of course, many get the item correct, irrespective of whether they are good or poor pupils.

It should be remembered that these indices are only valid for a particular examinee population at a particular point in time, and that the indices are NOT absolute values characteristic of the items. However, if the examinee population does not change significantly from year to year, we can use the data from a previous year to guide us in the design of new items and new tests.

APPENDIX II

Some Notes on Structured and Free Response Questions

by J.C. Mathews

1. In the past it has been usual to classify examination questions into two types, objective and essay. To the first there is only one correct answer. To the second an answer consists of an extended piece of writing, marked, with varying degrees of flexibility, from a predetermined marking scheme.

Objective questions usually include a number of suggested answers of which the candidates must pick one. They are objective in that the marking of the answers, and to some extent their purpose, are not so dependent on the judgment of individuals as are other forms of question. However, there is a considerable subjective element in the writing and selection of these questions, and it would be better if they were called fixed response questions.

While acknowledging that fixed response questions have brought to examining an exactness which other forms of question have lacked, it is now generally recognised that the sole use of them in school examinations may not be desirable. In class, teachers allow their pupils to respond freely to open-ended questions. Examinations - if they are to encourage desirable teaching situations, should do likewise.

The conventional essay-type question suffers from two defects. While apparently allowing a free response from candidates, in fact it does not; because, in an attempt to achieve reliable marking, examiners usually predetermine what they want candidates to say and use a highly structured marking scheme. The difficulty to the candidate is that this structuring is not apparent to him and he is left to guess what it is that the examiner requires him to say. Over the years, of course, teachers come to know what the examiner expects, and guide their pupils accordingly; but this leads to a rigidity in examinations, and a resistance to change by both examiner and teacher.

It seems self-evident that, if an examiner predetermines the structure and material of the answer he expects to a question, he should make the structure explicit in the question, and not leave the determination of this structure by the candidate to either guessing or custom. If, on the other hand, an examiner wants a candidate to give a genuine free response to a question, he must phrase the question sufficiently loosely to allow this response, and he must mark the answer in a manner in which he does not predetermine its structure in detail.

It appears, therefore, that three main types of question are necessary:

1.1 Fixed response questions. These may be in various forms, but they have one thing in common: the correct answer is fixed and stated and candidates have to choose it from other suggested answers which are also stated.

1.2 Structured questions. In these the examiner has predetermined in detail the structure of the answers he expects, and he has structured each question into parts which correspond to the expected answers. The parts of the question are usually open ended.

1.3 Free response questions. In these the examiner has not predetermined the answers which he expects. He will give general guidance to those who mark the answers, but not a detailed marking scheme.

2. Structured questions

2.1 The main features of a structured question should resemble those of a lesson. Modern teaching requires pupils to work on information rather than simply to recall it, so a structured question has a stem in which information is given to the candidates.

2.2 The stem. The data in the stem of a question should be concise so that candidates are not overburdened with reading and comprehension. If the questions are to test higher abilities than the ability to recall, the data should be partly unfamiliar to candidates or familiar material presented in an unfamiliar way. The stem should indicate what the overall purpose of the question is, so that as a candidate works through the parts he knows the general direction in which he is moving.

The information in the stem can be in various forms:

2.21 A description of a laboratory experiment - often with diagrams.

2.22 A description of an industrial process - often with diagrams.

2.23 Data about the behaviour of substances: descriptive, tabular, graphical, diagrammatic and other forms.

2.24 A passage from scientific literature.

2.3 The questions. These are best conceived by imagining the questions one would ask of a class faced with the information given in the stem of the question.

2.31 Most of the questions comprising a set should be open-ended. They should require short answers, and be so phrased that the examiner can predetermine what the answers should be and the candidates are in no doubt about what the examiner expects.

2.32 Within a set, questions should start with easy ones and increase in difficulty.

2.33 Questions should not depend too much on a candidate having given a correct answer to a question earlier in a set.

2.34 An appropriate space should be left on the question paper so that the candidates know the maximum amount that they are expected to write.

2.35 Writers of questions should have in mind the educational objective of each part of a question in a set.

2.36 Writers should try to look at the stem and the questions from the pupils' point of view, and a marking scheme should be considered and written at the same time that the questions are being written. The marking should be detailed and an appropriate mark allocated for each point which is expected in the answers, but there may have to be some marking by impression, particularly if a short comment or explanation is demanded. In the impression marking the marks should not exceed 2 or 3 for each part. With a finely structured set of questions and marking scheme, a high degree of reliability of marking can be obtained.

3. Free response questions

Free response questions can be asked about any part of a course of study, but they are particularly useful for those parts of a course in which there is a choice of topics or in which there is a large choice of material within a topic.

3.1 The questions should allow the maximum freedom of choice of material and expression to the candidates, and rarely should the candidates be restricted to particular substances, reactions or processes. It is in these questions that candidates who have read outside their basic course and those who are capable of divergent and original thought will be able to gain merit.

3.2 The examiner should look for general qualities rather than particular knowledge and candidates should be told what the examiner is looking for. If, for example, credit will be given for the ability to write grammatically and clearly, candidates should be told this either in the question or in a general rubric.

3.3 Many of the questions will be short, but there is no reason why this should always be so. A good free response question could be one in which a lot of data is given, the candidates being asked to make various generalisations and comments on the data in the form of an extended answer.

3.4 The mark scheme should be in the form of general guidance rather than a finely structured scheme. Marking is best done by impression on a six-point scale: 5, 4, 3, 2, 1, 0, the process being one of grading rather than mark allocation. Multiple marking is desirable.

4. A general comment to question writers

Remember that what you write will greatly influence what is taught and how it is taught. If, for example, practical work has a central position in a course it must have a central position in an examination.

Appendix A

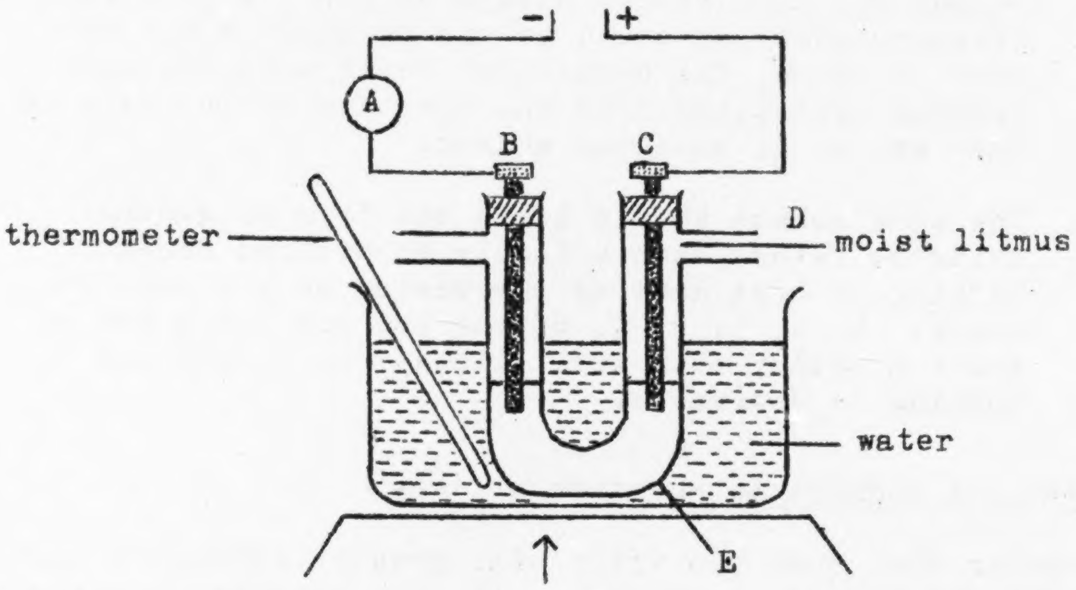
CEYLON 'O' LEVEL UNITS

IONIZATION I AND IONIZATION II

Structured Questions

Questions 1 - 9

This group of questions is based on an investigation into the effect of an electric current on a concentrated solution of lead chloride in water, at a temperature between 90° and 95°C. The apparatus drawn below was used.



1. What general name is given to the two pieces of apparatus B and C?

....

2. What particular name would be given to the piece of apparatus C?

....

3. Name a suitable material of which both B and C could be made.

....

4. What purpose is served by the piece of moist litmus paper at D?

....

5. Suggest a reason for heating the U-tube by surrounding it with hot water rather than heating it directly at E.

....

....

....

6. What would you expect to be formed on B?

....

7. When the temperature of the solution was allowed to fall, the current decreased. Suggest two reasons for this.

(i) ....

....

(ii) ....

....

8. This electrolysis resembles a change which takes place when a piece of zinc is placed into a solution of lead chloride. State and explain one similarity.

....

....

....

....

....

9. This electrolysis also differs from the change which takes place when a piece of zinc is placed into a solution of lead chloride. State and explain one difference.

....

....

....

....

....

....

....

Questions 10 - 17

This group of questions requires the following data:

Solubility in water at 20°C of:

silver chloride ($AgCl$) = 0.00000062 g-formula (moles)/litre
silver sulphate (Ag_2SO_4) = 0.0018 g-formula/litre
silver fluoride (AgF) = 1.4 g-formula/litre
silver bromide = 0.00000011 g-formula/litre

10. Arrange these four substances in the order of their solubility (mol/litre) putting the least soluble first.

....

11. Where would you expect to put silver IODIDE in this list?

....

12. What would you expect to observe (if anything) if a solution of potassium bromide were added to a saturated solution of silver sulphate?

....

13. What would you expect to observe (if anything) if a solution of sulphuric acid were added to a saturated solution of silver chloride?

....

14. What is the concentration of silver ions (g-ions/litre) in a saturated solution of silver chloride?

....

15. What is the concentration of silver ions (g-ions/litre) in a saturated solution of silver sulphate?

....

16. Write an equation representing the equilibrium which exists when a saturated solution of silver bromide is in contact with solid silver bromide.

....

17. How would you demonstrate that a saturated solution of silver bromide did contain ions? Do not give practical details.

....

....

....

....

....

....

....

MARKING SCHEME FOR STRUCTURED QUESTIONS

1. Electrode (1)

2. Anode (1)

3. Carbon is the most likely, but allow any inert electrode material (1)

4. To test for chlorine (1)

5. To get easier control and more uniform heating. Mark by impression (2-1-0)

6. Lead (1)

7. Fall in concentration of ions (1)
   Fall in mobility of ions (1)

8. For any reasonable similarity (the most likely is the formation of lead) (1). Explanation should include the fact that both the zinc and the cathode provide electrons for the change: Pb^{2+} (aq) + $2e^- \longrightarrow$ Pb (c). Mark by impression (2-1-0)

9. For any reasonable difference (for example, the formation of zinc ions in solution) (1). Mark explanation by impression (2-1-0)

Total for this set 15

10. Bromide, chloride, sulphate, fluoride (1)

11. In front of bromide (1)

12. Precipitate of silver bromide (1)

13. Nothing (1)

14. 0.00000062 (1)

15. 2 x 0.0018 (1)

16. $AgBr(c) \rightleftharpoons Ag^+(aq) + Br^-(aq)$ (2) 1 mark for minor error

17. Electrical conductivity (1) compared with that of pure water (1)

Total for this set 10

SPECIFICATION FOR THE STRUCTURED QUESTIONS

Weighting is according to the marks allotted.

Question	Activity	Ability
1.	B	1
2.	B	1
3.	B	3
4.	B	2
5.	B	3
	B	4
6.	A	2
7.	C	3
	C	4
8.	A	4
	E	3
	E	3
9.	A	4
	E	3
	E	3
10.	D	2
11.	C	4
12.	B	3
13.	B	3
14.	D	2
15.	D	3
16.	D	3
	E	3
17.	B	3
	B	3

Activities:
A - CHANGING MATERIALS
B - PRACTICAL TECHNIQUES
C - PERCEIVING PATTERNS
D - MEASURING AND CALCULATING
E - USING CONCEPTS

Abilities:
1. KNOWLEDGE
2. UNDERSTANDING
3. APPLICATION
4. ANALYSIS
5. SYNTHESIS